Published in final edited form as:

J Econom. 2016 February 1; 191(2): 276–292. doi:10.1016/j.jeconom.2015.12.001.

Dynamic Treatment Effects

James J. Heckman,

University of Chicago, and the American Bar Foundation

John Eric Humphries, and

University of Chicago

Gregory Veramendi

Arizona State University

Abstract

This paper develops robust models for estimating and interpreting treatment effects arising from both ordered and unordered multistage decision problems. Identification is secured through instrumental variables and/or conditional independence (matching) assumptions. We decompose treatment effects into direct effects and continuation values associated with moving to the next stage of a decision problem. Using our framework, we decompose the IV estimator, showing that IV generally does not estimate economically interpretable or policy relevant parameters in prototypical dynamic discrete choice models, unless policy variables are instruments. Continuation values are an empirically important component of estimated total treatment effects of education. We use our analysis to estimate the components of what LATE estimates in a dynamic discrete choice model.

Keywords

choice theory; dynamic treatment effects; factor models; marginal treatment effects; regret; conditional independence; matching on mismeasured variables; instrumental variables; ordered choice models; unordered choice models

James Heckman, Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, Phone: 773-702-0634, jjh@uchicago.edu

John Eric Humphries, Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, Phone: 773-980-6575, johneric@uchicago.edu

Gregory Veramendi, Department of Economics, Arizona State University, 501 East Orange Street, CPCOM 412A, Tempe, AZ 85287-9801, Phone: 480-965-0894, gregory.veramendi@asu.edu

*We thank the editor, Esfandiar Massoumi, and two anonymous reviewers for helpful comments. This research was supported in part by the American Bar Foundation, the Pritzker Children's Initiative, NIH grants NICHD R37HD065072, NICHD R01HD54702, and NIA R24AG048081. The views expressed in this paper are those of the authors and not necessarily those of the funders or persons named here or the official views of the National Institutes of Health. The Web Appendix for this paper is <https://heckman.uchicago.edu/dyn-treat-effects>.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 Introduction

This paper develops a robust empirical framework for estimating treatment effects arising from multistage decision models and interpreting them using economic theory. The bulk of the empirical treatment effect literature estimates models for binary choices. Although there is research on treatment effects for multiple choices¹, little analysis has been done for models of dynamic treatment effects.² Yet much economics is about dynamic choices and their consequences.

Figure 1 presents a schematic for one simple multistage choice model we analyze. It is in the form of the ordered choice model that is implicit in the multi-stage analysis of Angrist and Imbens (1995).³ The stages could correspond to a sequence of educational choices. All agents start at stage “0.” Some transit to “1,” while others stay at “0” forever, and some of those who go to 1 stop there while others go on, etc. At each stage, agents update their information sets and decide whether or not to transit to the next stage. Associated with each final stage is a set of potential outcomes. After we analyze this simple ordered model, we analyze a more general unordered model.

A large econometric literature analyzes dynamic discrete choice.⁴ These models tightly parameterize agent decision rules using the Bellman equation and generally rely on strong functional form assumptions and computationally intensive methods to secure their estimates.⁵ The complexity of the computational methods employed often makes replication and sensitivity analyses with these models difficult. In many applications, the sources of identification are not clear.⁶ Rust (1994) shows that an important class of these models is nonparametrically non-identified.⁷ Blevins (2014) shows how adding continuous state variables aids in securing nonparametric identification.

This paper steps back from the structural literature and presents a computationally tractable yet economically interpretable framework that enables analysts to identify their models, to conduct sensitivity analyses and to test some of the key assumptions maintained in the dynamic discrete choice literature. At the same time, it extends the treatment effect literature by considering dynamic treatment regimes, and by introducing choice-theoretic underpinnings.

This paper builds on and extends the literature on the Marginal Treatment Effect (MTE) that unifies the treatment effect literature with economics without imposing strong functional form assumptions or assumptions about specific decision rules adopted by agents.⁸

¹See Heckman and Vytlačil (2007b).

²See, however, Murphy (2003) and Heckman and Navarro (2007). This paper builds on the analysis reported in the latter reference. Angrist and Imbens (1995) develop a statistical model for multiple treatment effects that can be applied to a dynamic choice setting. Their paper identifies a LATE for an ordered choice model. (See Vytlačil, 2006a,b.) We identify a more general range of parameters for both ordered and unordered models.

³See Vytlačil (2006a,b) who establishes this.

⁴See e.g. Rust (1994), Keane and Wolpin (1997), Cameron and Heckman (1998), Heckman and Navarro (2007), Cunha, Heckman, and Navarro (2007), Aguirregabiria (2010), and Blevins (2014).

⁵See Adda and Cooper (2003) and Keane, Todd, and Wolpin (2011).

⁶However, Taber (2000) and Blevins (2014) present nonparametric identification analyses under separability and conditional independence assumptions.

⁷Magnac and Thesmar (2002) clarify and extend his analysis.

⁸See Heckman and Vytlačil, (1999); (2005); (2007a); (2007b); (2010) and Eisenhauer, Heckman, and Vytlačil (2015).

Empirical applications of the MTE focus on binary choices. Extensions of IV to ordered choice models and more general unordered multistate choice models demonstrate the need to incorporate explicit choice theory into analyses in order to identify a range of economically interpretable treatment effects beyond LATE parameters.⁹ Previous analyses based on MTE and LATE rely exclusively on instrumental variables to identify parameters.

This paper extends the literature by using conditional independence assumptions as well as instrumental variable assumptions as possible sources of identification. Conditional independence assumptions are used extensively in the dynamic discrete choice literature (see, e.g., Rust, 1994 and Blevins, 2014) and the matching literature (see, e.g., Rosenbaum and Rubin, 1983). They are especially well motivated if analysts have rich data on the determinants of choices. We extend the matching literature by considering models with mismeasured match variables on which analysts have multiple measurements.

This paper also builds on previous analyses of dynamic treatment effects presented in Cunha, Heckman, and Navarro (2007) and Heckman and Navarro (2007). We implement and extend the ordered choice model of Cunha, Heckman, and Navarro (2007) to allow for general stage-specific cost and preference shocks associated with learning as well as dynamically inconsistent preferences (see, e.g., Laibson, 2003) and for an unordered choice model. Using our model we can test for the empirical relevance of *ex-post* regret. We extend the work of Heckman and Navarro (2007) by building a more explicit economic framework of dynamic treatment effects which decomposes them into direct effects and continuation values for both ordered and unordered models, by linking their work to the matching literature, and by drawing on recent advances in identifying factor models.

Our analysis links treatment effect models to state space models where analysts can proxy unobservables. The proxies can be the true values of variables measured with error as in factor models (See Schennach, 2013).

This paper proceeds in the following way. The next section presents models for ordered and unordered choice, and distinguishes the approach pursued in this paper from that pursued in the previous literature. Section 3 defines dynamic treatment effects and their decomposition into direct effects and continuation values as well as a variety of other economically interpretable treatment parameters. Section 4 discusses some identification criteria for these models. Section 5 uses these models to interpret what IV estimates. Section 6 presents empirical estimates of the causal effects of schooling on earnings, decomposing them into direct and continuation value components. It tests and rejects some key maintained assumptions in dynamic discrete choice theory, and compares estimates of economically interpretable parameters with LATE. We use our analysis to estimate what LATE can and cannot estimate in dynamic discrete choice models. When possible, we resolve LATE into economically interpretable components. Section 7 concludes.

⁹Heckman, Urzúa, and Vytlačil (2006, 2008) and Heckman and Pinto (2015a).

2 Models for Ordered and Unordered Dynamic Discrete Choice and Associated Outcomes

This paper develops a multistage ordered sequential choice model with transitions at the nodes shown in Figure 1. For specificity, it is useful to think of the nodes as corresponding to specific schooling levels through which individuals can transit or at which they can stop. An unordered model is analyzed after we analyze the ordered model depicted in Figure 1. Let $\mathcal{J}$ denote an *ordered* set of possible terminal states. At each node there are only two possible choices: remain at j or transit to $j+1$. $D_j = 0$ if a person at j does not stop there and goes on to $j+1$. $D_j = 1$ if the person stops at j . $D_j \in \mathcal{D}$, the set of possible transition decisions that can be taken by the individual over the decision horizon. Let $\mathcal{S} = \{0, \dots, s\}$ denote the finite and bounded set of stopping states with $S = s$ if the agent stops at $s \in \mathcal{S}$, so $D_s = 1$. Define $\bar{s}$ as the highest attainable element in $\mathcal{S}$. We assume that the environment is time-stationary and decisions are irreversible.¹⁰

$Q_j = 1$ indicates that an agent *gets to* decision node j . $Q_j = 0$ if the person never gets there. The history of nodes visited by an agent can be described by the collection of the Q_j such that $Q_j = 1$. Observe that $D_s = 1$ and $D_{s-1} = 0$ are equivalent to $S = s$ and $D_s = 1$ if $D_j = 0, \forall j < s$.¹¹

2.1 A Sequential Decision Model

The decision process at each node is assumed to be characterized by an index threshold-crossing property:

$$D_j = \begin{cases} 0 & \text{if } I_j \geq 0, \\ 1 & \text{otherwise,} \end{cases} \quad j \in \mathcal{J} = \{0, \dots, \bar{s}-1\} \quad \text{for } Q_j = 1, \quad j \in \{0, \dots, \bar{s}-1\} \quad (1)$$

where I_j is the *perceived* value at node j of going on to $j+1$ for a person at node j . To ensure consistent notation, we define $Q_0 = 1$. The requirement $Q_j = 1$ ensures that agents are able to make the transition from j to $j+1$. Unlike the literature on discrete dynamic choice, we do not take a position on the precise content of the information sets and preferences governing agent choices at different nodes or the exact decision rules used. In principle, agents could make irrational choices and their choices could be governed by behavioral anomalies. Agents face stage-specific cost and/or preference shocks that they may or may not anticipate before they reach the stage. Thus, this model does not rule out dynamic inconsistency as featured in behavioral economics (Laibson, 2003), or in adolescent psychology (Steinberg, 2014) where agent preferences evolve, and *ex-post*, agents regret irreversible decisions made at previous stages.

Associated with each final state $s \in \mathcal{S} = \{0, \dots, s\}$ is a set of K_s potential outcomes for each agent with indices $k \in \mathcal{K}_s$. Define $\tilde{Y}_s^k$ as latent variables that map into potential outcomes Y_s^k :

¹⁰This model is also analyzed in Cunha, Heckman, and Navarro (2007) and in Heckman and Navarro (2007).

¹¹For notational convenience, we assign $D_j = 0$ for all $j > s$.

$$Y_s^k = \begin{cases} \tilde{Y}_s^k & \text{if } Y_s^k \text{ is continuous,} \\ \mathbf{1}(\tilde{Y}_s^k \geq 0) & \text{if } Y_s^k \text{ is a binary outcome,} \end{cases} \quad k \in \mathcal{K}_s, \quad s \in \mathcal{S}. \quad (2)$$

Using the switching regression framework of Quandt (1958, 1972) the observed outcome Y^k for a k common to $\bigcap_{s \in \mathcal{S}} \mathcal{K}_s$ is

$$Y^k = \sum_{s \in \mathcal{S}} D_s Y_s^k, \quad k \in \mathcal{K}_s. \quad (3)$$

2.2 Parameterizations of the Decision Rules and Potential Out-comes for Final States

Following a well-established tradition in the discrete choice literature, we represent I_j using a separable model:

$$I_j = \phi_j(\underbrace{\mathbf{Z}}_{\text{Observed by analyst}}) + \underbrace{\overbrace{\boldsymbol{\theta}' \boldsymbol{\alpha}_j - \nu_j}}^{\text{Unobserved by analyst}}, \quad j \in \{0, \dots, \bar{s}-1\} \quad (4)$$

$-\eta_j$

where $\phi_j(\mathbf{Z})$ is a function of a vector of variables, $\mathbf{Z}$, observed by the analyst, components of which determine the transition decisions of the agent at different stages and $\boldsymbol{\theta}$ is a finite dimensional vector of unobserved (by the economist) endowments. ν_j is an idiosyncratic transition-specific innovation. It is notationally useful to consolidate the unobservables into scalar $\eta_j = -[\boldsymbol{\theta}' \boldsymbol{\alpha}_j - \nu_j]$. Separability is a cornerstone of the dynamic discrete choice literature.¹² A separable representation of choice equations also underlies the LATE literature (Vytlačil, 2002, 2006a,b).

With this representation we can test implications of more tightly specified versions of dynamic discrete choice models. Endowments $\boldsymbol{\theta}$ are not directly observed by the econometrician but are proxied by a measurement system discussed below. Array the $\nu_j, j \in \mathcal{J}$, into a vector $\boldsymbol{\nu} = (\nu_0, \nu_1, \dots, \nu_{s-1})$ and the η_j into $\boldsymbol{\eta} = (\eta_0, \dots, \eta_{s-1})$.

Outcomes are also represented by a separable model:

¹²See, e.g., Heckman (1981), Cameron and Heckman (1987, 2001), Eckstein and Wolpin (1989), Geweke and Keane (2001), Heckman and Navarro (2007), Arcidiacono and Ellickson (2011), and Blevins (2014).

$$\tilde{Y}_s^k = \underbrace{\tau_s^k(X)}_{\substack{\text{observed} \\ \text{by analyst}}} + \underbrace{\theta' \alpha_s^k + \omega_s^k}_{U_s^k}, \quad k \in \mathcal{K}_s, \quad s \in \mathcal{S}, \quad (5)$$

unobserved
by analyst

where $\tau_s^k(X)$ is a function of a vector of observed determinants of outcomes and θ is the vector of unobserved endowments. The Z can include all variables in X . When instrumental variable methods are used to identify components of the model, it is assumed that there are some variables in Z not in X .

ω_s^k represents an idiosyncratic error term for outcome k in state s . Array the ω_s^k into a vector $\omega_s = (\omega_s^1, \dots, \omega_s^{K_s})$. Notationally, it is useful to consolidate the unobservables of (5) into a term $U_s^k = \theta' \alpha_s^k + \omega_s^k$. Array the U_s^k into vector $U_s = (U_s^1, \dots, U_s^{K_s})$ and array the U_s into $U = (U_0, \dots, U_S)$.

θ is assumed to capture all sources of dependence across transitions apart from X and Z . It plays an important role in our analysis and links the structural model to the matching literature. Along with the observed variables, it generates the dependence between choices and outcomes.

Letting “ $\perp$ ” denote statistical independence, we assume that *conditional on X*

$$\nu_j \perp \nu_l \quad \phi l \neq j \quad l, j \in \{0, \dots, \bar{s}-1\} \quad (\text{A-1a})$$

$$\omega_s^k \perp \omega_{s'}^k \quad \phi s \neq s' \quad \phi k \quad (\text{A-1b})$$

$$\omega_s \perp \nu \quad \phi s \in \mathcal{S} \quad (\text{A-1c})$$

$$\theta \perp Z \quad (\text{A-1d})$$

$$(\omega_s, \nu) \perp (\theta, Z) \quad \phi s \in \mathcal{S}, \quad (\text{A-1e})$$

Assumption (A-1a) maintains independence of the shocks affecting transitions; (A-1b) independence of shocks across all states; (A-1c) independence of the shocks to transitions and the outcomes; (A-1d) independence of θ with respect to the observables; and (A-1e) independence of the shocks and the factors with the observed variables. Versions of assumptions (A-1d) and (A-1e) play fundamental roles in the structural dynamic discrete choice literature. For example, the widely-used “types” assumption of Keane and Wolpin (1997) postulates conditional independence between choices and outcomes conditional on types (θ) that operate through the initial conditions of their model.

2.3 A Measurement System for Unobserved Factors θ

A substantial body of the literature estimating causal effects develops strategies for eliminating the effect of θ in producing spurious relationships between the final states selected and outcomes.¹³ A principal approach to identification adopted in this paper is to model θ and its relationship to the choice equations and associated outcomes. We adjoin a measurement system to the system of choice and outcome equations, as in state space models. We proxy θ to identify the interpretable sources of omitted variable bias and to determine how unobservables mediate causal effects. This enables us to apply our analysis to an emerging literature in economics and personality psychology that investigates the role of cognitive and noncognitive skills in shaping choices and final outcomes.¹⁴

If θ were observed, one could condition on $(\theta, \mathbf{X}, \mathbf{Z})$ and identify selection-bias-free estimates of a variety of causal effects. This is equivalent to matching on $\theta, \mathbf{X}$, and $\mathbf{Z}$.¹⁵ Both matching and the dynamic discrete choice literature assume that conditional on $\theta, \mathbf{X}$, and $\mathbf{Z}$, outcomes and choices are statistically independent.

The structural literature treats the θ as nuisance variables, invokes conditional independence assumptions, and integrates out the unobservables using random effect procedures (see e.g., Keane and Wolpin, 1997; Rust, 1994; Adda and Cooper, 2003 and Blevins, 2014). In this paper we proxy θ with multiple measurements and correct for the effects of measurement error on the proxy.

Let $\mathbf{M}$ be a vector of N_M measurements on θ . They may consist of lagged or future values of the outcome variables or additional measurements.¹⁶ The system of equations determining $\mathbf{M}$ is:

$$\mathbf{M} = \Phi(\mathbf{X}, \theta, \mathbf{e}), \quad (6)$$

where $\mathbf{X}$ are observed variables, θ are the endowments and

$$\mathbf{M} = \begin{pmatrix} M_1 \\ \vdots \\ M_{N_M} \end{pmatrix} = \begin{pmatrix} \Phi_1(\mathbf{X}, \theta, e_1) \\ \vdots \\ \Phi_{N_M}(\mathbf{X}, \theta, e_{N_M}) \end{pmatrix}$$

where we array the e_j into $\mathbf{e} = (e_1, \dots, e_{N_M})$. We assume in addition to the previous assumptions that *conditional on $\mathbf{X}$*

¹³See e.g. Heckman (2008) for a review of alternative approaches in the literature.

¹⁴The recent literature establishes that both cognitive and noncognitive skills can be shaped by interventions and are effective avenues for social policy. See, e.g., Borghans, Duckworth, Heckman, and ter Weel (2008); Almlund, Duckworth, Heckman, and Kautz (2011); Heckman, Humphries, and Kautz (2014a); Heckman, Pinto, and Savelyev (2013); Heckman and Kautz (2014). See Heckman, Humphries, and Kautz (2014b) for such an application.

¹⁵Matching is a version of selection on observables. See Carneiro, Hansen, and Heckman, 2003, and Abbring and Heckman, 2007.

See also Heckman and Vytalacil (2007a,b) and Heckman, Schennach, and Williams (2011).

¹⁶See, e.g., Abbring and Heckman (2007); Schennach, White, and Chalak (2012).

$$e_j \perp e_l, \quad j \neq l, \quad j, l \in \{1, \dots, N_M\} \quad (\text{A-1g})$$

$$\text{and } e \perp (\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\omega}). \quad (\text{A-1h})$$

We assume that $(\mathbf{e}, \boldsymbol{\nu}, \boldsymbol{\omega})$ conditional on $\mathbf{X}$ is a vector of absolutely continuous variables. Let $(\bar{\mathbf{e}}, \bar{\boldsymbol{\nu}}, \bar{\boldsymbol{\omega}})$ and $(\underline{\mathbf{e}}, \underline{\boldsymbol{\nu}}, \underline{\boldsymbol{\omega}})$ be the vector of upper and lower limits of $(\mathbf{e}, \boldsymbol{\nu}, \boldsymbol{\omega})$, respectively.

2.4 An Unordered Dynamic Discrete Choice Model

The ordered model just analyzed is traditional in the literature on the economic returns to schooling that assumes that some measure of years of schooling captures exposure to education.¹⁷ Yet there are many forms of education that cannot be directly summarized by a years of schooling equivalent.¹⁸ Examples are GED certificates,¹⁹ post-school training programs, vocational degrees, and the like. There is no natural order on these measures of educational attainment.

To focus on a simple case that describes the data analyzed in this paper, we analyze a model with the GED. Adding it to Figure 1 produces a new path from the node “drops out of high school” described in Figure 1. Interpreting node “0” as being in high school but not graduating, the path “remain at 0” corresponds to being a dropout from high school. However, the GED creates a new path from the dropout state: get a GED. We assume that GED attainment is a terminal state. In truth, many GEDs attempt college although few graduate.²⁰ In the samples we analyze there are too few GEDs going on to college to estimate a stable model for that transition.

Research shows that there is no natural years of schooling to assign to GEDs.²¹ Figure 2 adds a branch to the lower portion of Figure 1. The upper portion is as before except we now give specificity to the nodes and transitions. The decision tree depicted in Figure 2 is more typical of discrete dynamic decision processes estimated in the literature.²²

To conserve on notation, for agents for whom $D_0 = 0$, we define the set of attainable nodes as before. However, we now change slightly the interpretation of $D_0 = 1$. Dropping out opens a new branch of the decision tree. For $D_0 = 1$, we define the attainable sets as $\{0, G\}$. Thus, a person may remain a dropout or may exam certify. $Q_G = 1$ if the agent drops out of high school and confronts the GED option. Thus, in the upper branch, $(D_0 = 0)$, and the $D_j, j = 1$, are as before. In the lower branch $(D_0 = 1)$, agents can terminate as a dropout $(D_0 = 1, D_G = 1)$ with outcome Y_0 or as a dropout who exam certifies $(D_0 = 1, D_G = 0)$ with outcome Y_G .

¹⁷The Mincer model (1974) is the leading example.

¹⁸However, equivalents might be formed using outcomes in some outcome metrics such as earnings or occupation. These would give tautological explanations of outcomes.

¹⁹The GED is an exam whose proponents claim that successful examinees are the equivalents of high school graduates. For strong evidence to the contrary, see Heckman, Humphries, and Kautz (2014a).

²⁰Heckman, Humphries, and Kautz (2014a).

²¹Heckman, Humphries, and Kautz (2014a).

²²See e.g., Rust, 1994.

As before, outcomes at each terminal node may be discrete or continuous (see Equation 2). The terminal state space $\mathcal{S}$ is amended to include G , i.e., $\mathcal{S} = \{G, 0, \dots, s\}$. For outcome k , the Quandt switching regression representation for observed Y^k for a k common to $\bigcap_{s \in \mathcal{S}} \mathcal{H}_s$ is

$$Y^k = \left(\sum_{\mathcal{S} \setminus \{0, G\}} D_s Y_s^k \right) (1 - D_0) + \left(Y_0^k D_G + Y_G^k (1 - D_G) \right) D_0. \quad (7)$$

The decision model is now characterized by two branches. The origin node (drop out of school) is characterized as before:

$$D_0 = \begin{cases} 0 & \text{if } I_0 \geq 0 \\ 1 & \text{otherwise.} \end{cases}$$

For the upper path, given $D_0 = 0$,

$$D_j = \begin{cases} 0 & \text{if } I_j \geq 0, j=1, \dots, J \\ 1 & \text{otherwise} \end{cases}.$$

For the lower path, given $D_0 = 1$,

$$D_G = \begin{cases} 0 & \text{if } I_G \geq 0 \\ 1 & \text{otherwise} \end{cases}.$$

The parameterization of the index functions is as before with an obvious change of notation.

3 Dynamic Treatment Effects

The treatment effects defined in this paper take into account the direct effect of transiting to the next node in a decision tree, plus the benefits associated with the options opened up by the additional choices made possible by such transitions. Under autonomy, different treatment effects can be identified by fixing treatment assignment variables for different subpopulations.

Dynamic discrete choice models greatly facilitate the interpretation of intertemporal choices and their consequences. As an example, consider a dynamic human capital model analyzed by Keane and Wolpin (1997). Assume risk-neutral agents who have a finite choice set with N alternatives over a finite decision horizon ($\underline{a}$). Let $B_n(a) = 1$ if alternative n is chosen at age a and zero otherwise. Let $R_n(a)$ be the current flow reward at age a from alternative n .²³ The current reward per period at any age a is

²³Alternative sets may vary

$$R(a) = \sum_{n=1}^N \underbrace{R_n(a)}_{\substack{\text{per} \\ \text{period} \\ \text{reward} \\ \text{from} \\ \text{choice } n}} \underbrace{B_n(a)}_{\substack{\text{choice} \\ \text{indicator}}} .$$

Denote an individual's state at age a by $\mathbf{H}(a)$. Assume a discount factor δ . The value function is

$$V(\mathbf{H}(a), a) = \max_{B_n(t) \in \mathcal{B}(a)} E \left[\sum_{t=a}^{\bar{a}} \delta^{t-a} \sum_{n=1}^N R_n(t) B_n(t) | \mathbf{H}(a) \right]$$

where $\mathcal{B}(a)$ is the set of feasible current and future choices at age a . The alternative-specific functions, $V_n(\mathbf{H}(a), a)$ can be written as

$$V_n(\mathbf{H}(a), a) = R_n(\mathbf{H}(a), a) + \underbrace{\delta E[V(\mathbf{H}(a+1), a+1) | \mathbf{H}(a), B_n(a)=1]}_{\text{Continuation Value}}$$

for $a < \bar{a}$, where $V_n(\mathbf{H}(a), a) = R_n(\mathbf{H}(a), a)$, the reward in state n at age a for a person with history $\mathbf{H}(a)$. The decision rule is $B_n(a) = 1$ if $n = \arg\max_{j \in \{1, \dots, N\}} \{V_j(\mathbf{H}(a), a)\}$; $B_n(a) = 0$ otherwise.

Fully specified dynamic discrete choice models postulate agent preferences, constraints, and information sets, and can recover continuation values associated with each choice as well as option values that arise from moving through decision trees like those of Figures 1 and 2.²⁴ These benefits come at a price and many empirical economists reject the strong assumptions invoked in using dynamic economic models.²⁵

This paper takes a more agnostic and data-sensitive approach. We estimate a dynamic treatment effect model that captures some essential features of dynamic discrete choice models, but does not impose specific functional forms, decision models, and assumptions about distributions of unobservables.

In the rest of this section, we analyze treatment effects associated with the ordered model of Figure 1 and note the modifications required, if any, for the more general model of Figure 2. While it is possible to define a variety of treatment effects consistent with the model, many would be empirically implausible.

²⁴See, e.g., Eisenhauer, Heckman, and Mosso (2015) for estimates of a structural model of schooling with option values and continuation values.

²⁵See, e.g., Angrist and Pischke (2010).

Many empirical economists would not find estimates of the effect of fixing $D_j = 0$ if $Q_j = 0$ credible (i.e., the person for whom we fix $D_j = 0$ is not at the decision node to take the transition).²⁶ In this spirit, we define treatment effects associated with fixing $D_j = 0$ conditioning on $Q_j = 1$. (For the lower branch of the unordered model we require $Q_G = 1$ in analyzing the effect of fixing D_G .)

The person-specific treatment effect T_j^k for outcome $k \in \bigcap_{s \in \mathcal{S}} \mathcal{K}_s$ for an individual selected from the population $Q_j = 1$ with characteristics $\mathbf{X} = \mathbf{x}$, $\mathbf{Z} = \mathbf{z}$, $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$, making a decision at node j between going on to $j+1$ or stopping at j is the difference for the individual between the individual's outcomes under the two actions:

$$T_j^k[Y^k | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, \boldsymbol{\theta} = \bar{\boldsymbol{\theta}}] := (Y^k | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, \boldsymbol{\theta} = \bar{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 0) - (Y^k | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, \boldsymbol{\theta} = \bar{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 1), \quad (8)$$

where the random variable $(Y^k | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, \boldsymbol{\theta} = \bar{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 0)$ is the value of Y^k at node j for a person with characteristics $\mathbf{X} = \mathbf{x}$, $\mathbf{Z} = \mathbf{z}$, $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$ from the population that attains node j (or higher), $Q_j = 1$, and for whom we fix $D_j = 0$ so that agents are forced to go on to the next node. Random variable $(Y^k | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, \boldsymbol{\theta} = \bar{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 1)$ is defined for the same individual but forces the person with these characteristics not to transit to the next node. (For the unordered model we account for the conditioning on branch membership, i.e., $D_0 = 0$ or $D_0 = 1$, by conditioning on Q_j .)

The person-specific treatment effect can be decomposed into two components: the *Direct Effect* of going from j to $j+1$: $DE_j^k = Y_{j+1}^k - Y_j^k$, and the *Continuation Value* of going beyond $j+1$:

$$C_{j+1}^k = \sum_{r=1}^{\bar{s}-(j+1)} \left[\prod_{l=1}^r (1 - D_{j+l}) \right] (Y_{j+r+1}^k - Y_{j+r}^k). \quad (9)$$

In the unordered model, the continuation value for the upper branch is defined as before (conditioning on $D_0 = 0$) and for the lower branch it is defined for the attainable set $\{0, G\}$. Essentially, G is the only option available to a high school dropout in our model.

At the individual level, the *Total Effect* of fixing $D_j = 0$ on Y^k is decomposed into

$$T_j^k = DE_j^k + C_{j+1}^k. \quad (10)$$

The associated population level average treatment effect conditional on $Q_j = 1$ is

$$ATE_j^k := \int \dots \int E[T_j^k(Y^k | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, \boldsymbol{\theta} = \bar{\boldsymbol{\theta}})] dF_{\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}, \bar{\boldsymbol{\theta}} | Q_j = 1). \quad (11)$$

Integrating over $\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}$, conditioning on $Q_j = 1$, the population continuation value at $j+1$ is

²⁶The distinction between *fixing* and *conditioning* traces back to Haavelmo (1943). White and Chalak (2009) use the terminology “setting” for the same notation. For a recent analysis of this crucial distinction, see Heckman and Pinto (2015b).

$$E_{\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}}(C_{j+1}^k) = E_{\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}} \left[\sum_{l=j+1}^{\bar{s}-1} \left\{ E(Y_{l+1}^k - Y_l^k | \mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}, \boldsymbol{\theta}=\bar{\boldsymbol{\theta}}, Q_{l+1}=1, \text{Fix } Q_{j+1}) \right. \right. \\ \left. \left. \cdot Pr(Q_{l+1}=1 | \text{Fix } Q_{j+1}=1, \mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}, \boldsymbol{\theta}=\bar{\boldsymbol{\theta}}, Q_j=1) \right\} | Q_j=1 \right]. \quad (12)$$

where $Q_s = 1$ if $S = s$.

We can also specify population distributions of total effects as in Heckman, Smith, and Clements (1997):²⁷

$$Pr(T_j^k < t_j^k | \mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}, \boldsymbol{\theta}=\bar{\boldsymbol{\theta}}, Q_j=1) \quad (13)$$

with population distribution counterpart

$$E_{\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}} \left[Pr(T_j^k < t_j^k | Q_j=1) \right] \quad (14)$$

which can be decomposed into the distribution of direct effects and continuation values. (The modifications for the unordered case require that we define these terms over the admissible options available for $D_0 = 1$ or $D_0 = 0$.)

We thus approximate the treatment effects that could be obtained from a fully specified dynamic discrete choice model. However, we abstract from age-specific rewards and consider only stage-specific rewards (which may be discounted return streams associated with each choice).²⁸ Because we do not specify or attempt to identify choice-node-specific agent information sets, we only analyze *ex-post* treatment effects, and cannot identify option values associated with choices.

Since our framework does not impose specific decision rules, it is consistent with irrationality, regret, and mistakes in decision-making associated with agent maturation and acquisition of information. In the context of a schooling model, with our framework we can identify the proportion of agents with *ex-post* regret about their educational decisions (if they are maximizing the present value of earnings) and can identify the characteristics of the population with regret. We can also test predictions of fully specified dynamic discrete choice models, i.e., that expected later-stage costs and rewards should affect agent transitions at earlier stages. Our framework thus enables analysts to examine the validity of key assumptions maintained in the structural discrete choice literature.

Many different treatment effects can be defined depending on the populations and margins targeted. We now present two that are not standard in the statistical treatment effect literature but that answer interesting economic and policy questions.

²⁷See Abbring and Heckman (2007) for a review of the literature.

²⁸See, e.g., Heckman and Navarro (2007).

3.1 Average Marginal Treatment Effects

The Average Marginal Treatment Effect (AMTE)²⁹ is the average effect of transiting to the next stage for individuals at the margin of indifference between the two stages:

$$AMTE_j^k := \int \dots \int E \left[T_j^k \left(Y^k | \mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}, \boldsymbol{\theta}=\bar{\boldsymbol{\theta}}, |I_j| \leq \varepsilon \right) \right] dF_{\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}, \bar{\boldsymbol{\theta}} | Q_j=1, |I_j| \leq \varepsilon), \quad (15)$$

where ε is an arbitrarily small neighborhood around the margin of indifference. AMTE defines causal effects at well-defined and empirically identified margins of choice. It is the proper measure of marginal gross benefits for evaluating the gains of moving from one stage of the decision node to the next for those at that margin of choice, and in general is distinct from LATE.³⁰ The population distribution counterpart of AMTE is defined over the set of agents for whom $|I_j| \leq \varepsilon$, which can be generated from our model:

$Pr(T_j^k < t_j^k | Q_j=1, |I_j| \leq \varepsilon)$.³¹ definition, with modification for the branch of Figure 2 analyzed, also applies to the unordered case.

3.2 Policy Relevant Treatment Effects

The policy relevant treatment effect (PRTE) is the average treatment effect for those induced to change their choices in response to a particular policy intervention. Let $Y^k(p)$ be the aggregate outcome under policy p for outcome k . Let $S(p)$ be the final state selected by an agent under policy p . The policy relevant treatment effect from implementing policy p compared to policy p' for outcome k is:

$$PRTE_{p,p'}^k := \int \dots \int E(Y^k(p) - Y^k(p') | \mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}, \boldsymbol{\theta}=\bar{\boldsymbol{\theta}}, S(p) \neq S(p')) dF_{\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}, \bar{\boldsymbol{\theta}} | S(p) \neq S(p')), \quad (16)$$

where $S(p) \neq S(p')$ denotes the set of the characteristics of people for whom attained states differ under the two policies. It is different from AMTE because different agents affected by the policy can be at different margins of choice. PRTE is often confused with LATE. In general, they are different unless the policy change coincides with the instrument used to define LATE. The differences between the two can be substantial.³²

Proceeding in a similar fashion, we can define conventional treatment effects, e.g., ATE, TT, TUT defined for different populations. Again, population distribution counterparts can be defined for the distribution of the gains for any subpopulation with positive probability. The modifications required for the unordered case are straightforward. One needs to specify the options available for the branches of Figure 2 on which agents operate.

4 Identification

The structural approach to econometrics estimates the full set of equations and distributions of the unobservables specified in Equations (1) – (6) under the assumptions postulated in

²⁹See Carneiro, Heckman, and Vytlacil (2010, 2011).

³⁰See, e.g., Heckman and Vytlacil (2007a) and Carneiro, Heckman, and Vytlacil (2010).

³¹Because we keep the interval ε fixed and finite, we avoid the non-uniqueness of limits for the limits of AMTE discussed in Carneiro, Heckman, and Vytlacil (2010), although there are different estimates for different ε .

³²See Carneiro, Heckman, and Vytlacil (2011) for example.

Section 2. Access to these equations enables analysts to generate all of the treatment effects discussed in Section 3 and more. This approach has great appeal. At issue is the credibility of the resulting estimates. The appeal of the treatment effect approach is that identification of parameters demands fewer and often more transparent assumptions. The downside of the approach is that the treatment effects so generated often have unknown economic or policy relevance. The approach in this paper offers a compromise between these two approaches.

This section gives sufficient conditions for identification of the full model of Section 2. It then considers some simpler criteria useful for identifying some of the treatment effects defined in Section 3.

4.1 Identification of the Full Structural Model: The Ordered Case

One approach to identification of the full structural model builds on the analyses of Heckman and Navarro (2007) as refined in Abbring and Heckman (2007).³³ These authors exploit variation in the $\mathbf{Z}$ of Equation 1 across transitions as well as restrictions on $\phi_j(\mathbf{Z})$. Such variation can arise from exclusion restrictions in IV analysis or from functional form restrictions across determinants of transition equations. After developing this approach, we develop an approach based on conditional independence assumptions as used in matching and in the structural literature on dynamic discrete choice.

Heckman and Navarro (2007) analyze the identifiability of a single spell discrete time duration model with an associated vector of potential outcomes at each stopping time. The time periods in their model correspond to the stages in Figure 1. Their decision model corresponds to our Equation (1).

Their Theorem 1 does not invoke any factor structure assumption, and works directly with the η_j . It applies to the linear-in-parameters specification of the choice index

$$I_j(\mathbf{Z}) = \mathbf{Z}_j \boldsymbol{\gamma}_j - \eta_j \quad j \in \mathcal{J} = \{0, \dots, \bar{s}-1\}. \quad (17)$$

We consider a more general specification after we consider the analysis of a linear-in-parameters model which serves to fix ideas. The relevant ($\mathbf{Z}$) components are assumed to satisfy Assumptions (A-1d) and (A-1e). Assume access to a random sample of data on $\mathbf{D}$ and $\mathbf{Z}$ and assume no censoring of transition histories on individuals. It is convenient to introduce the following notation. Let $\mathbf{Z} = (\mathbf{Z}_0, \dots, \mathbf{Z}_{s-1})$ and $\boldsymbol{\eta} = (\eta_0, \dots, \eta_{s-1})$. Define $\mathbf{Z}^j = (\mathbf{Z}_0, \dots, \mathbf{Z}_j)$ and $\boldsymbol{\eta}^j = (\eta_0, \dots, \eta_j)$ for $j \in \{0, \dots, s-1\}$. Analogously, define $\boldsymbol{\gamma}^j = (\gamma_0, \dots, \gamma_j)$ and $\mathbf{D}^j = (\mathbf{D}_0, \dots, \mathbf{D}_j)$ for $j \in \{0, \dots, s-1\}$. Let $\boldsymbol{\eta} = (\eta_0, \dots, \eta_{s-1})$ and $\underline{\boldsymbol{\eta}} = (\underline{\eta}_0, \dots, \underline{\eta}_{s-1})$ denote the upper and lower limits of the support of $\boldsymbol{\eta}$, with $\boldsymbol{\eta}^j$ and $\underline{\boldsymbol{\eta}}^j$ defined comparably for $\boldsymbol{\eta}^j$.

Theorem 4.1: *Under the following assumptions, conditional on $\mathbf{X}$,*

- i. *(i) $\mathbf{Z}^j \perp\!\!\!\perp \boldsymbol{\eta}^j$ (implied by (A-1d) and (A-1e)).*

³³Taber (2000) presents a different approach for a two stage version of this model.

- ii. η is absolutely continuous with respect to Lebesgue measure defined on the support

$$\prod_{l=0}^j (\underline{\eta}_l, \bar{\eta}_l)$$

- iii. **Full rank:** For all $l \in \{0, \dots, j\}$, $\mathbf{Z}_l$ is a W_l dimensional vector. There exists no proper linear subspace of $\mathcal{R}^{W_l}$ having probability 1 under $F_{\mathbf{Z}_l}$, the cdf of $\mathbf{Z}_l$. There exists a $\mathbf{g}^j = (\mathbf{g}_0, \dots, \mathbf{g}_j)$, a vector of constants, such that for almost all $\mathbf{g}^j = (\mathbf{g}_0, \dots,$

$$\mathbf{g}_j) \prod_{l=0}^j (\underline{\eta}_l, \bar{\eta}_l) \text{ with } \mathbf{g}^j \text{ } j \text{ component-wise, there exists no proper linear subspace of } \mathcal{R}^{W_0} \times \mathcal{R}^{W_1} \times \dots \times \mathcal{R}^{W_j} \text{ having probability 1 under } F_{\mathbf{Z}_0 \gamma_0 > g_0, \dots, \mathbf{Z}_{j-1} \gamma_{j-1} > g_{j-1}}.$$

- iv. **Inclusion of supports:**

$$\text{Supp}(\mathbf{Z}_j \gamma_j | \mathbf{Z}_0 \gamma_0 = g_0, \dots, \mathbf{Z}_{j-1} \gamma_{j-1} = g_{j-1}) \supseteq (\underline{\eta}_j, \bar{\eta}_j) \quad (18)$$

where the boundary points $(\underline{\eta}_j, \bar{\eta}_j)$ are not functions of γ_j . Then, F_{η^j} and η^j are identified up to a scale and location normalization.

Proof: Heckman and Navarro (2007), *Theorem 1*.

The independent variation requirements of Assumption (iii) and the inclusion of supports condition of Assumption (iv) can be stringent in applications. In empirical work these conditions are sometimes difficult to satisfy although, as we show in Section E of the Web Appendix, they are satisfied in our data. Corollary 1 of Heckman and Navarro (2007) ensures identification (up to the usual scale and location normalizations) without any exclusion restrictions, provided that the coefficients arrayed in $\gamma^j = (\gamma_0, \dots, \gamma_j)$ are linearly independent and there are at least j continuous (with respect to Lebesgue measure) components of each of the $\mathbf{Z}^j$.

Heckman and Navarro (2007) extend Theorem 4.1 in two important ways. First, they consider a class of general functional forms for $I_j(\mathbf{Z}) = \phi_j(\mathbf{Z}) - \eta_j$ that satisfy the Matzkin (1994) conditions.^{34, 35} The Matzkin class includes the linear in the parameters model (equation 17) as a special case. Second, to the discrete choice model of Equations (1) and (4) they adjoin a vector of outcomes $\mathbf{Y}_j, j \in \mathcal{J}$ which may be discrete or continuous, but if discrete, are generated by dichotomizations of continuous latent variables.³⁶

They prove the following theorem, which we adapt to our case.

Theorem 4.2: Assume data on $(\mathbf{Y}_j, \mathbf{D}^j, \mathbf{X}, \mathbf{Z})$ given $Q_j = 1$ from a random sample across individuals. Assume that $\phi_j(\mathbf{Z})$ belong to the class of Matzkin functions and array them in vector $\phi(\mathbf{Z}) = (\phi_0(\mathbf{Z}), \dots, \phi_j(\mathbf{Z})), j \in \{0, \dots, s-1\}$. Assume further that:

³⁴For convenience of the reader, we reproduce these conditions in Web Appendix A.1.

³⁵They can relax the independence assumption ($\mathbf{Z} \perp \eta^j$) provided that in each $\mathbf{Z}_j$, one regressor satisfies the special regressor assumptions of Lewbel (2000). See the Web Appendix of Heckman and Navarro (2007).

³⁶This can be readily extended to variables with multiple discrete outcomes using Theorem 1 in Heckman and Navarro (2007).

- i. (U_j, η_j) are absolutely continuous random variables with zero means, finite variances and support $\text{Supp}(U_j) \times \text{Supp}(\eta_j)$ with upper and lower limits $(\bar{U}_j, \bar{\eta}_j)$ and $(\underline{U}_j, \underline{\eta}_j)$ respectively with similar conditions holding for each subvector.
- ii. $(U_j, \eta_j) \perp\!\!\!\perp (X, Z), j \in \mathcal{J}$, an implication of Equations (A-1e) and (A-1d)
- iii. $\text{Supp}(\tau_j(X), \phi_j(Z)) = \text{Supp}(\tau_j(X)) \times \prod_{l=1}^j \text{Supp}(\phi_l(Z))$
- iv. $\text{Supp}(\phi_j(Z)) \supseteq \text{Supp}(\eta_j)$

Then $\tau_j(X)$, $\phi_j(Z)$, $F_{\eta_j|U_j}$ are identified up to scale if the Matzkin functions for the discrete components are specified up to scale, and are exactly identified if a specific normalization is used for those components.

Proof: Heckman and Navarro (2007), Theorem 2.

Theorem 4.2 does not identify the joint distribution of $Y_0, \dots, Y_j$ because analysts observe only one of these vector outcomes for any person. Observe that we do not require strict exclusion restrictions in the arguments of the treatment choice equation to identify the model. We require independent variation of arguments that might be achieved by instrumental variable exclusion conditions but that can also be obtained by other functional restrictions such as those previously discussed for the linear model of Theorem 4.1.³⁷ Note further that it is possible to identify the $\tau_j(X)$ (up to constants) without any limit set argument.³⁸

As a by-product of Theorem 4.2, under autonomy (i.e. structural invariance), we can construct the distributions of Y_j for a variety of counterfactual histories up through j

$$\Pr(Y_j \leq y_j | D_0 = d_0, \dots, D_j = d_j) \quad (19)$$

for $d_l \in \{0, 1\}$. The condition $Q_j = 1$ in (1) imposes the requirement that

$$0 \leq \sum_{l=0}^j D_l \leq 1 \quad \text{for all } j \in \{0, \dots, \bar{j}\}. \quad (\text{C-1})$$

Otherwise, event sequences with probabilities greater than one could be assumed to characterize the counterfactual world.

Note that the counterfactuals that are identified by fixing D_j at different values have an asymmetric aspect. One can generate the Y_j distributions for persons who terminate their transitions in state j or before. Without further structure, one cannot generate the distributions of the Y_{j+k} , $k > 0$ for people who do not transit beyond j .

³⁷For examples, see Cameron and Heckman (1998) and Carneiro, Hansen, and Heckman (2003).

³⁸See the discussion in Heckman and Navarro (2007).

The source of this asymmetry is the generality of the model. At each node j , agents acquire a new random variable η_j which can have arbitrary dependence with Y_j and $Y_{j'}$ for all j and j' . From Theorem 4.2, we can identify the dependence between η_j and $Y_{j'}$ if $j = j'$. We cannot identify the dependence between η_j and $Y_{j'}$ for $j < j'$ without imposing further structure on the unobservables, which we do in the next subsection. Thus, for example, we can identify the distribution of college outcomes for high school graduates who do not go on to college and can compare these with their distribution of high school outcomes. Thus from Theorem 4.2 we can identify the parameter “treatment on the untreated.” However, we cannot identify the distribution of high school outcomes for college graduates (*e.g.* treatment on the treated parameters) without imposing further structure.³⁹

Note that Theorem 4.2 does not invoke the factor structure assumptions postulated in Section 3. Nonetheless from Theorem 4.2, we can construct all marginal and joint probabilities of $D = (D_0, \dots, D_{s-1})$. Observe that $D_j = 0 \Rightarrow Q_{j+1} = 1$. We can construct $Pr(Q_j = 1 | Q_j = 1)$ in expression (12) for all l as well as $E(Y_{j+1} - Y_j | Q_j = 1, Q_j = 1)$. Hence, over the relevant supports we can identify $E(C_{j+1})$. However without further restrictions, we cannot identify

$$E(Y_j | D_{j+l} = d_{j+l}, D_j = d_j, \dots, D_0 = d_0) \quad l \geq 1. \quad (20)$$

Hence, we *cannot* identify treatment on the treated parameters or (without further assumptions) any of the distributional treatment effects.⁴⁰ We can get distributional treatment effects under rank invariance of outcomes (across the j) or invoking some other known relationship among quantiles of outcomes across successive j . Invoking a factor structure assumption also facilitates identification of these counterfactuals and provides an alternative way to identify versions of the model of Section 2.

4.1.1 The Unordered Case—The analysis for the unordered case is a straightforward extension of the analysis of Heckman and Navarro (2007) and for the sake of brevity is omitted. It entails applying the conditions of Theorems 1 and 2 to the different decision branches of Figure 2. We can identify $Pr(D_0 = 1 | Z)$ under the conditions of Theorems 1 and 2. We can then apply the conditions of Theorems 1 and 2 which use identification at infinity to achieve limit sets in which $Pr(D_0 = 1) = 1$ and $Pr(D_0 = 0) = 1$, analyze the two branches separately and proceed as before.

4.2 Identification Using a Factor Structure

Factor models make possible identification of a broader class of counterfactuals (see Carneiro, Hansen, and Heckman, 2003; Abbring and Heckman, 2007, and Heckman and Navarro, 2007). They exploit the dependence across choices and outcomes generated by θ .

³⁹It is straightforward to establish identification (for all admissible histories up to j) of (a) Average treatment effects for all final state attainment levels, $s \in S$; (b) Average effects of treatment on the untreated; (c) The AMTE; (d) The Total Effect, the Average Direct Effect, and the Average Continuation Values for counterfactuals satisfying (C-1). Consider $E(Y_j | D_0 = d_0, \dots, D_j = d_j)$. Under the full support condition of Theorem 4.2, we can construct the entire sequence of counterfactuals $E(Y_j | D_0 = d_0, \dots, D_{j-1} = d_{j-1}), \dots, E(Y_j | D_0 = d_0), E(Y_j)$. We can do this for all $j \in \mathcal{J}$. Hence, we can form all of the listed mean treatment effects by taking suitable limit operations. For the sake of brevity, we do not provide the proofs, which exploit the linear properties of means and the continuity of the

⁴⁰See Heckman, Smith, and Clements (1997).

With sufficient numbers of measurements $\mathbf{M}$, one can identify the joint distribution of θ and bring that to the table to identify the structural model of Section 2. Knowledge of θ secured from $\mathbf{M}$ can enable analysts to identify the joint distributions of counterfactuals and choices $(\mathbf{D}, \mathbf{Y})$ even though they do not observe all of the components of $\mathbf{Y}$ across final schooling states for anyone. Thus knowledge of the dependence across measurements can compensate for the lack of knowledge of the dependence across counterfactuals.⁴¹

This knowledge provides an alternative to the exclusive use of instrumental variable models for identifying structural or treatment effect parameters. The dynamic discrete choice literature invokes conditional independence assumptions (see, e.g., Rust, 1994, Magnac and Thesmar, 2002, or Blevins, 2014), as does the matching literature.

Factor structure models enhance the interpretability of estimated treatment effects. When the factors can be related to interpretable measurements, analysts can interpret the sources of the unobservables that give rise to treatment effects and how they are related to the determinants of the choice equations. In the context of analyses of the effects of education on earnings, factor models facilitate investigations of the sources of ability bias and its empirical importance. Dynamic factor structure models have been used to identify the sources of skill development. (See Cunha, Heckman, and Schennach, 2010 and Heckman, Pinto, and Savelyev, 2013). We consider a variety of cases, starting with the simplest. All of the analysis applies, in a straightforward way, to both the ordered and unordered cases with a suitable modification of notation.

4.2.1 Known θ —Under the assumptions of Section 2, if researchers know θ , they can condition on it and identify all of the treatment effects previously defined, including treatment on the treated and distributional treatment effects, over the available support of $\mathbf{X}$, $\mathbf{Z}$, θ . This is true for both the ordered and unordered model. If these conditions apply, additive separability can be relaxed and standard results in matching and nonparametric econometrics can be invoked to identify parameters under well-known conditions.⁴²

4.2.2 Unknown But Proxied θ —The more challenging case arises when θ is measured with error but can be proxied by measurements which may include lagged or future values of outcomes and choices.⁴³ Schennach (2013) provides a valuable survey of the literature.

Identification of factor models is inherently controversial. Without an associated measurement system, factor models lack interpretability. In our approach, the adjoined measurement system $\mathbf{M}$ facilitates interpretability. But, as is well-known from linear factor models, and is true in the general nonlinear system of equation (6), many alternative sets of identifying conditions can rationalize a given set of measurements due to classical rotation and scaling problems.

⁴¹Abbring and Heckman provide a variety of simple examples showing how measurements and panel data on outcomes with each state, $s \in \{0, \dots, s\}$ can identify the full distribution of $(\mathbf{D}, \mathbf{Y})$.

⁴²See, e.g., Pagan and Ullah (1999).

⁴³See Abbring and Heckman (2007).

If the goal of a study is to identify the counterfactual choice and outcome equations, but not to interpret the effects of θ , it is not necessary to solve these rotation and scale problems (see Heckman, Schennach, and Williams, 2011, revised 2014). Any measure-preserving transformation of the random variables that span θ can be used as conditioning variables. One can bypass discussions of identifiability of the measurement equations (6).

One particularly simple case, applied in Heckman, Pinto, and Savelyev (2013), Gensowski (2014), and Kautz and Zanoni (2015), extracts factor scores from a linear measurement system $\mathbf{M}$, and uses the estimated factor scores as Bartlett (1937) regressors. It corrects for measurement error in estimating the θ using the analyses of Skrondal and Laake (2001). Measurement error arises from the fact that Bartlett scores are only estimates of θ . Unlike traditional measurement error problems, from the factor system it is possible to estimate the distribution of the measurement error. Armed with this knowledge, it is possible to correct for the measurement error in estimating treatment effects and structural equations. The analysis is particularly simple when the outcome and choice equations are linear functions of the factors and the measurements are continuous. With linear measurements and nonlinear choice and outcome equations, a variety of methods for analyzing nonlinear models with mismeasured conditioning variables but known distributions of measurement error are available to estimate model parameters.

Using the analyses of Hu and Schennach (2008), Cunha, Heckman, and Schennach (2010), and Freyberger (2014), it is possible to identify the distributions of factors even when the measurement equations (6) are nonlinear, and to apply the methods just discussed. As noted in the survey by Schennach (2013), there has been considerable progress in identifying factor models using information on higher moments and using nonlinearities intrinsic to the general factor model.⁴⁴ A full account of the specific application of these various methods is outside the scope of this paper.

The analysis of Freyberger (2014) is immediately applicable to this paper and we draw on it. He analyzes the nonparametric identification and estimation of a measurement system $\mathbf{M}$ (in our notation) with a linear factor structure:

$$M_j = \Phi_j(\mathbf{X}) + \lambda_j' \theta + e_j$$

$j = 1, \dots, N_M$. Under his conditions, which for the sake of brevity we do not present here, it is possible to identify nonparametrically the $\Phi_j(\mathbf{X})$, the λ_j , the distribution of θ and the distribution of $\mathbf{e} = (e_1, \dots, e_m)$ over the population support.

Armed with this result and the analysis of Theorem 4.2, and in particular knowledge of the distribution of $(\eta_0, \dots, \eta_{s-1})$ and recalling that

$$\eta_l = \alpha_l' \theta - \nu_l, \quad l \in \{0, \dots, s-1\}$$

⁴⁴See, e.g., Bonhomme and Robin (2009).

and arraying the α_l , $l = 0, \dots, s - 1$ into an $s \times s$ matrix A

$$\Sigma_{\eta} = A' \Sigma_{\theta} A + \Sigma_{\nu}$$

where Σ_{ν} is an $s \times s$ diagonal matrix of uniquenesses and Σ_{θ} is the variance-covariance matrix of θ .

With knowledge of Σ_{η} we can identify Σ_{ν} . From knowledge of Σ_{θ} we can identify A (up to standard normalizations).⁴⁵ Since we know the distribution of θ , we know the distribution of $\alpha_l' \theta$ and by deconvolution we can identify the distribution of ν . Applying the same type of argument to the outcome equations we achieve identification of the full model and we can construct the full joint distribution of counterfactuals.⁴⁶

5 Understanding What Instrumental Variables Estimate

This section uses our model to interpret what IV estimates. IV has a prominent place in the recent literature in applied economics. A variety of “causal effects” are offered in that literature, often without any clear economic motivation. We relate our analysis to the MTE of Heckman and Vytlacil (1999, 2005) which in the binary case links a theoretical IV literature with economically interpretable and policy-relevant treatment effects. We consider IV in both the ordered and unordered model. The simplicity and tractability of the MTE in the case of binary choices is compromised because of the complex conditioning that arises from dynamic choices. In this section, we analyze what is estimated by: (1) local instrumental variables, (2) Wald IV, and (3) instrumental variables with a continuous instrument.

To focus ideas, and to anchor our analysis in the conventional literature, it is useful to have as a touchstone a prototypical model in labor economics that has been a focus of much of the recent literature on instrumental variables.⁴⁷ The bulk of the empirical literature on the returns to schooling is based on linear instrumental variables estimates of a version of the Mincer equation for person i :⁴⁸

$$\underbrace{\ln y_i}_{\text{log earnings}} = \alpha_i + \underbrace{\beta_i}_{\text{“rate of return”}} \underbrace{S_i}_{\text{schooling}}. \quad (21)$$

⁴⁵See, e.g., Anderson and Rubin (1956).

⁴⁶Indeed, drawing on the analyses of Abbring and Heckman (2007) and Freyberger (2014), it appears possible to identify the full joint distribution of policy counterfactuals using a sufficiently rich set of outcomes and choice data without using any measurement system. We do not develop that analysis in this paper. We note in addition that using the application of the analyses of Madansky (1964), Chamberlain (1977) and Pudney (1982) by Freyberger we can relax the requirements that $Z \perp\!\!\!\perp \theta | X$.

⁴⁷See, e.g., Angrist and Imbens (1995) and Angrist and Pischke (2009).

⁴⁸See Card (1999); Heckman, Lochner, and Todd (2006); Lochner (2011); Mincer (1974) and McMahon (2014).

β_i is the causal effect of schooling on earnings and is often called a “rate of return” to schooling.⁴⁹ It is assumed that, contrary to the model underlying Figure 2, education is captured by a scalar years of schooling variable. When β_i is the same constant for all i and y_i and S_i are measured without error, the only source of bias in estimating the model is the dependence between a_i and S_i (ability bias). When β_i is correlated with S_i , there is also selection on gains.

This specification combines years of schooling into a linear aggregator and ignores any dynamics associated with continuation values. When $\beta_i = \beta$ or $\beta_i \perp S_i$, one valid instrument identifies the (mean) return across all schooling levels. In a companion paper (Heckman, Humphries, and Veramendi, 2015), we test for and reject the linear specification (21) within the upper branch of Figure 2. There is selection on gains (β_i) at higher schooling levels but not at lower levels.

We consider what IV estimates when the dynamic sequential structural model analyzed in this paper generates choices and outcomes.⁵⁰ Dynamic models are known to pose challenges for standard instrumental variable procedures. If agents (even partially) anticipate instruments that affect future choice decisions, the anticipated values of those instruments are arguments of the choice equations preceding those decisions through the Bellman relationship, compromising standard IV exclusion and exogeneity restrictions.⁵¹ We investigate another source of IV exclusion restrictions arising from the effects of lagged instruments in affecting the conditioning sets of agents making current decisions. Instruments from previous transitions determine the conditions leading up to any current choice set.

In the notation of our model, for the ordered case $Y = \sum_{j=0}^{\bar{s}} D_j Y_j$ and

$$\begin{aligned} E(Y|Z) &= \left(\sum_{j=0}^{\bar{s}} E(D_j Y_j | Z) \right) \\ &= \sum_{j=0}^{\bar{s}} E(Y_j | D_j = 1, Z) E(D_j | Z) \end{aligned} \quad (22)$$

where, definitionally, $D_j = 0 \forall j > \bar{s}$ implies $D_{\bar{s}} = 1$. (The corresponding term for the unordered case applies the law of iterated expectations to Equation 7.)

To relate our general model to the conventional literature on the returns to schooling based on (21), let q_j be the years of schooling assigned to level j . q_j need not equal j or be any affine function of the j , but $q_{j-1} < q_j$ for all $j \in \{0, \bar{s}\}$.⁵² In this notation, in the ordered model, schooling is

⁴⁹It is under the special conditions discussed in Heckman, Lochner, and Todd (2006).

⁵⁰Heckman and Urzúa (2010) consider this question in the context of a static multinomial discrete choice model. New phenomena arise in the model investigated in this paper.

⁵¹See, e.g., Hansen and Sargent (1980). As shown there, innovations to the instruments remain valid exclusion restrictions.

⁵²Thus, persons enrolling in college but not completing it spend roughly 2.4 years there (so $q_3 = 14.4$), while college graduates spend (typically) 4 years (so $q_4 = 16$).

$$S = \sum_{j=0}^{\bar{s}} q_j D_j$$

where $Pr(S = q_j | \mathbf{Z}) = Pr(D_j = 1 | \mathbf{Z})$ and

$$E(S | \mathbf{Z}) = \sum_{j=0}^{\bar{s}} q_j E(D_j | \mathbf{Z}). \quad (23)$$

Explicit expressions for these component terms are useful for understanding what IV

estimates in the ordered model. Because terminal states are mutually exclusive, $\sum_{j=0}^{\bar{s}} D_j = 1$. Substituting for D_0 in the preceding expressions, we obtain some useful representations:

$$\mathbf{Y} = \mathbf{Y}_0 + \sum_{j=1}^{\bar{s}} (\mathbf{Y}_j - \mathbf{Y}_{j-1}) \sum_{l=j}^{\bar{s}} D_l \quad (24)$$

and

$$S = q_0 + \sum_{j=1}^{\bar{s}} (q_j - q_{j-1}) \sum_{l=j}^{\bar{s}} D_l. \quad (25)$$

Applying the model of Section 2 and keeping the $\mathbf{X}$ implicit, we obtain the following expression:

$$\begin{aligned} E(\mathbf{Y} | \mathbf{Z} = \mathbf{z}) &= \int_{\mathcal{Y}_0} \mathbf{y}_0 \int_{\phi_0(\mathbf{z})}^{\bar{\eta}_0} f_{\mathbf{y}_0, \eta_0}(\mathbf{y}_0, \eta_0) d\mathbf{y}_0 d\eta_0 \\ &+ \sum_{j=1}^{\bar{s}-1} \int_{\mathcal{Y}_j} \mathbf{y}_j \int_{\phi_j(\mathbf{z})}^{\bar{\eta}_j} \cdots \int_{\bar{\eta}_0}^{\phi_0(\mathbf{z})} f_{\mathbf{y}_j, \eta_j, \dots, \eta_0}(\mathbf{y}_j, \eta_j, \dots, \eta_0) d\eta_j \dots d\eta_0 d\mathbf{y}_j \\ &+ \int_{\mathcal{Y}_{\bar{s}}} \mathbf{y}_{\bar{s}} \int_{\underline{\eta}_{\bar{s}-1}}^{\phi_{\bar{s}-1}(\mathbf{z})} \cdots \int_{\underline{\eta}_0}^{\phi_0(\mathbf{z})} f_{\mathbf{y}_{\bar{s}}, \eta_{\bar{s}-1}, \dots, \eta_0}(\mathbf{y}_{\bar{s}}, \eta_{\bar{s}-1}, \dots, \eta_0) d\eta_{\bar{s}-1} \dots d\eta_0 d\mathbf{y}_{\bar{s}}, \end{aligned} \quad (26)$$

where $\mathcal{Y}_j$ is the support of $\mathbf{Y}_j$ and

$$\begin{aligned} E(S | \mathbf{Z} = \mathbf{z}) &= q_0 \int_{\phi_0(\mathbf{z})}^{\bar{\eta}_0} f_{\eta_0}(\eta_0) d\eta_0 + \\ &+ \sum_{j=1}^{\bar{s}-1} q_j \int_{\phi_j(\mathbf{z})}^{\bar{\eta}_j} \cdots \int_{\underline{\eta}_0}^{\phi_0(\mathbf{z})} f_{\eta_j, \eta_{j-1}, \dots, \eta_0}(\eta_j, \dots, \eta_0) d\eta_j \dots d\eta_0 \\ &+ q_{\bar{s}} \int_{\underline{\eta}_{\bar{s}-1}}^{\phi_{\bar{s}-1}(\mathbf{z})} \cdots \int_{\underline{\eta}_0}^{\phi_0(\mathbf{z})} f_{\eta_{\bar{s}-1}, \dots, \eta_0}(\eta_{\bar{s}-1}, \dots, \eta_0) d\eta_{\bar{s}-1} \dots d\eta_0. \end{aligned}$$

Note that $\mathbf{Z}$ only enters through the limits of the integrals because of independence assumption (A-1e).

5.1 The Unordered Case

There is no counterpart to Equation (23) in the unordered case unless an arbitrary number is assigned to G . Counterparts to (24) and (25) are straightforward to derive when we condition on $D_0 = 0$ or $D_0 = 1$. For the sake of brevity we do not develop these expressions. The entire recent IV literature based on LATE implicitly assumes the ordered case of Figure 1. (See Vytlacil, 2006a,b and Heckman and Pinto, 2015a.)

5.2 Local Instrumental Variables

This section analyzes what is estimated by Local Instrumental Variables (LIV, see Heckman and Vytlacil, 1999, 2005) in the context of the ordered model. It is evident from Equation (26) that any LIV derived from it will depend on the path of instruments up to a transition node. Variation in an instrument that appears in multiple nodes will affect $\mathbf{Y}$ through multiple channels.

To begin, consider the simplified case where an instrument only affects the final schooling transition. Let $\mathbf{Z}$ be a continuous instrument that only affects the argument $\phi_{s-1}(\mathbf{Z})$ of transition $s-1$ (the final transition). Consider a scalar component of this instrument, say Z_{s-1}^n . Then under the conditions of Theorem 4.2 and assuming finite means, for an outcome m with $\phi_{s-1}(\mathbf{Z})$ continuously differentiable in Z_{s-1}^n

$$\frac{\partial E(\mathbf{Y}|\mathbf{Z})}{\partial Z_{s-1}^n} = \frac{\partial \phi_{s-1}(\mathbf{Z})}{\partial Z_{s-1}^n} \cdot E \left(\mathbf{Y}_{\bar{s}} - \mathbf{Y}_{\bar{s}-1} | \eta_{\bar{s}-1} = \phi_{s-1}(\mathbf{Z}), \phi_{\bar{s}-2}(\mathbf{Z}) \geq \eta_{\bar{s}-2}, \dots, \phi_0(\mathbf{Z}) \geq \eta_0 \right) \cdot f_{\eta_{\bar{s}-1}}(\phi_{s-1}(\mathbf{Z}) | \eta_{\bar{s}-2} < \phi_{\bar{s}-2}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \cdot Pr(\eta_{\bar{s}-2} < \phi_{\bar{s}-2}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \quad (27)$$

where

$$\begin{aligned} E(\mathbf{Y}_{\bar{s}} - \mathbf{Y}_{\bar{s}-1} | \eta_{\bar{s}-1} \\ &= \phi_{s-1}(\mathbf{Z}), \phi_{\bar{s}-2}(\mathbf{Z}) \geq \eta_{\bar{s}-2}, \dots, \phi_0(\mathbf{Z}) \geq \eta_0) = \text{MTE}_{\bar{s}, \bar{s}-1}(\eta_{\bar{s}-1} \\ &= \phi_{s-1}(\mathbf{Z}), D_{s-2}(\mathbf{Z}) \\ &= 0, \dots, D_0(\mathbf{Z}) = 0) \end{aligned}$$

is the mean gain for outcome $\mathbf{Y}$ from going from $s-1$ to s for a person with $\eta_{s-1} = \phi_{s-1}(\mathbf{Z})$ for whom $D_{s-2}(\mathbf{Z}) = 0, \dots, D_0(\mathbf{Z}) = 0$. It is the sequential dynamic counterpart to the MTE parameter introduced in Heckman and Vytlacil (1999, 2005).

The comparable expression for $E(S|\mathbf{Z})$ is

$$\frac{\partial E(S|\mathbf{Z})}{\partial Z_{s-1}^n} = \frac{\partial \phi_{s-1}(\mathbf{Z})}{\partial Z_{s-1}^n} \cdot (q_{\bar{s}} - q_{\bar{s}-1}) \quad (28)$$

$$\cdot f_{\eta_{s-1}}(\phi_{s-1}(\mathbf{Z})|\eta_{s-2} < \phi_{s-2}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \cdot Pr(\eta_{s-2} < \phi_{s-2}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \quad (29)$$

LIV(m, Z_{s-1}^n) is obtained as the ratio of the first expression over the second expression:

$$LIV = E \left(\mathbf{Y}_{\bar{s}} - \mathbf{Y}_{\bar{s}-1} | \eta_{\bar{s}-1} = \phi_{\bar{s}-1}(\mathbf{Z}), \phi_{\bar{s}-2}(\mathbf{Z}) \geq \eta_{\bar{s}-2}, \dots, \phi_0(\mathbf{Z}) \geq \eta_0 \right) \left(\frac{1}{q_{\bar{s}} - q_{\bar{s}-1}} \right),$$

assuming that (28) is nonzero. Observe that the entire history of instruments $s-1$ also enter this expression, unlike its counterpart in the binary case. Such conditioning on instruments is not standard in the IV literature and in its recent counterparts for binary choice models but is a feature of models with multiple choices.⁵³

This case is very special because it only considers variations that affect the final transition. We now develop the general case. In this case, even if an instrument only affects the argument of one transition (say the l^{th} , $\phi_l(\mathbf{Z})$), the instrument affects multiple outcomes as it percolates over the entire dynamic process.

To develop the general case, consider a continuous component Z^n that may appear in multiple arguments of the functions determining transitions $\phi_l(\mathbf{Z})$, $l \in \{0, \dots, s-1\}$. Straightforward calculations establish that

$$\frac{\partial E(\mathbf{Y}|\mathbf{Z})}{\partial Z^n} = \sum_{j=0}^{\bar{s}} \sum_{l=0}^{\min\{j, \bar{s}-1\}} \left(\frac{\partial \phi_l(\mathbf{Z})}{\partial Z^n} \right) \Omega(j, l) \quad (30)$$

where

$$\begin{aligned} \Omega(j, l) = & E(\mathbf{Y}_j | \phi_0(\mathbf{Z}) \geq \eta_0, \dots, \phi_{l-1}(\mathbf{Z}) \geq \eta_{l-1}, \eta_l = \phi_l(\mathbf{Z}), \phi_{l+1}(\mathbf{Z}) \geq \eta_{l+1}, \dots, \phi_j(\mathbf{Z}) < \eta_j) \\ & \times \{f_{\eta_l}(\phi_l(\mathbf{Z})|\eta_j > \phi_j(\mathbf{Z}), \eta_{j-1} < \phi_{j-1}(\mathbf{Z}), \dots, \eta_{l+1} < \phi_{l+1}(\mathbf{Z}), \eta_{l-1} < \phi_{l-1}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \\ & \cdot Pr(\eta_j > \phi_j(\mathbf{Z}), \eta_{j-1} < \phi_{j-1}(\mathbf{Z}), \dots, \eta_{l+1} < \phi_{l+1}(\mathbf{Z}), \eta_{l-1} < \phi_{l-1}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z}))\} \end{aligned} \quad (31)$$

for $j \in \{0, \dots, s-1\}$ and $l \leq j$, and where the term in braces on the right hand side of (31) (a weight) is the conditional density of η_l evaluated at $\phi_l(\mathbf{Z})$ given that $D_k = 0$, $k \in \{0, \dots, l-1, l+1, \dots, j-1\}$ and $D_j = 1$. This expression exists given the assumed absolute continuity of the $\eta_0, \dots, \eta_j$.⁵⁵

For $l = j$ and $j < s$

$$\Omega(j, j) = -E(\mathbf{Y}_j | \phi_0(\mathbf{Z}) > \eta_0, \dots, \phi_{j-1}(\mathbf{Z}) > \eta_{j-1}, \dots, \phi_j(\mathbf{Z}) = \eta_j) \quad (32)$$

⁵³See Heckman (2010).

⁵⁴Intermediate calculations useful for establishing this result are presented in Appendix A.2.

⁵⁵For the explicit expression, see Web Appendix A.2.

$$\times \left\{ f_{\eta_j}(\phi_j(\mathbf{Z}) | \eta_{j-1} < \phi_{j-1}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \cdot Pr(\eta_{j-1} < \phi_{j-1}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \right\} \quad (33)$$

The corresponding term for $j = \bar{s}$ is

$$\Omega(\bar{s}, l) = E(Y_{\bar{s}} | \phi_0(\mathbf{Z}) \geq \eta_0, \phi_l(\mathbf{Z}) = \eta_l, \dots, \phi_{\bar{s}-1}(\mathbf{Z}) \geq \eta_{\bar{s}-1}) \quad (34)$$

$$\times \left\{ f_{\eta_l}(\phi_l(\mathbf{Z}) | \eta_{\bar{s}-1} < \phi_{\bar{s}-1}(\mathbf{Z}), \dots, \eta_{l+1} < \phi_{l+1}(\mathbf{Z}), \eta_{l-1} < \phi_{l-1}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \cdot Pr(\eta_{\bar{s}-1} < \phi_{\bar{s}-1}(\mathbf{Z}), \dots, \eta_{l+1} < \phi_{l+1}(\mathbf{Z}), \eta_{l-1} < \phi_{l-1}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \right\}. \quad (35)$$

Equation (30) gives the sum of the marginal effects on Y of variation in the instrument as it affects selection into different stages of the decision tree in Figure 1. Such effects do not arise in the traditional binary choice LATE literature where the path leading up to a transition is not an issue. Notice that values of Z^n that affect the probability of being at transition l ($Pr(Q_l = 1 | \mathbf{Z})$) affect outcomes even if Z^n is not an argument of $\phi_l(\mathbf{Z})$.⁵⁶ Locally, the components of $\Omega(j, l)$ for which $\phi_l(\mathbf{Z})$ does not depend on Z^n do not affect the calculation.

By a parallel argument,

$$\frac{\partial E(S | \mathbf{Z})}{\partial Z^n} = \sum_{j=0}^{\bar{s}} \sum_{l=0}^{\min\{j, \bar{s}-1\}} q_j \left(\frac{\partial \phi_l(\mathbf{Z})}{\partial Z^n} \right) \tilde{\Omega}(j, l), \quad (36)$$

where $\tilde{\Omega}(j, l)$, $j = \bar{s}$ are:

$$f_{\eta_l}(\phi_l(\mathbf{Z}) | \eta_j > \phi_j(\mathbf{Z}), \eta_{j-1} < \phi_{j-1}(\mathbf{Z}), \dots, \eta_{l+1} < \phi_{l+1}(\mathbf{Z}), \eta_{l-1} < \phi_{l-1}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \cdot Pr(\eta_j > \phi_j(\mathbf{Z}), \eta_{j-1} < \phi_{j-1}(\mathbf{Z}), \dots, \eta_{l+1} < \phi_{l+1}(\mathbf{Z}), \eta_{l-1} < \phi_{l-1}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \quad (37)$$

where the corresponding term for $j = \bar{s}$ is

$$\tilde{\Omega}(\bar{s}, l) = f_{\eta_l}(\phi_l(\mathbf{Z}) | \eta_{j-1} < \phi_{j-1}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \cdot Pr(\eta_{j-1} < \phi_{j-1}(\mathbf{Z}), \dots, \eta_0 < \phi_0(\mathbf{Z})) \quad (38)$$

and the corresponding term for $j = 0$ is

$$f_{\eta_0}(\phi_0(\mathbf{Z})). \quad (39)$$

Collecting results

$$LIV(\mathbf{Z}^n) = \frac{\frac{\partial E(Y | \mathbf{Z})}{\partial Z^n}}{\frac{\partial E(S | \mathbf{Z})}{\partial Z^n}} \times \mathbf{1} \left[\left(\frac{\partial E(S | \mathbf{Z})}{\partial Z^n} \right) \neq 0 \right]. \quad (40)$$

⁵⁶Heckman, Urzúa, and Vytlačil (2008) report related results in a model of discrete choice. See also Heckman and Urzúa (2010).

Note that the Z^n can affect multiple margins. Note further that even if each $\phi_f(\mathbf{Z})$ is monotonic in Z^n , the term $\frac{\partial E(S|\mathbf{Z})}{\partial Z^n}$ can be 0 and can switch signs over the support of $\mathbf{Z}$. LIV is a weight average of derivatives of marginal valuation functions.

5.2.1 The Unordered Case for Local Instrumental Variables—There is no counterpart of the analysis of this section for the unordered case because there is no counterpart to (28). For an analysis of what IV identifies in this case, see Heckman and Urzúa (2010) and Heckman (2010).

5.3 Wald IV

This section decomposes the Wald estimator into underlying economically interpretable components. We find that in the ordered case, the estimator is a weighted average across many sub-populations and many margins. For the analysis of the unordered case, see Heckman and Urzúa (2010). The Angrist-Imbens analysis of multiple choices (1995) only applies to ordered models (See Vytlacil, 2006a,b and Heckman and Pinto, 2015a).

Under standard independence assumptions that are implied by the model of Section 2,⁵⁷ for an instrument vector $\mathbf{Z}$ set at two different values $\mathbf{Z} = \mathbf{z}_1$ and $\mathbf{Z} = \mathbf{z}_2$, assuming $S(\mathbf{z}_2) \geq S(\mathbf{z}_1)$, where $S(\mathbf{z}_k)$ is the random variable when $\mathbf{Z}$ is fixed at $\mathbf{z}_k$ and for specificity that $S(\mathbf{z}_2) > S(\mathbf{z}_1)$, the probability limit of the Wald IV for Equation (21) featured in the recent literature⁵⁸ for a scalar outcome in the ordered model is

$$Wald\ IV(z_1, z_2) = \frac{\sum_{j=1}^{\bar{s}} E(Y_j - Y_{j-1} | S(\mathbf{z}_2) \geq q_j > S(\mathbf{z}_1)) Pr(S(\mathbf{z}_2) \geq q_j > S(\mathbf{z}_1))}{\sum_{j=1}^{\bar{s}} (q_j - q_{j-1}) Pr(S(\mathbf{z}_2) \geq q_j > S(\mathbf{z}_1))}, \quad (41)$$

where $\frac{Y_j - Y_{j-1}}{q_j - q_{j-1}} = \beta$ for all $j = 1, \dots, \bar{s}$ in the Mincer model (21).⁵⁹ This estimand identifies the change in the aggregate outcome resulting from changes in schooling induced by the instrument.

$E(Y_j - Y_{j-1} | S(\mathbf{z}_2) \geq q_j > S(\mathbf{z}_1))$ is a “causal effect” and is often contrasted with clearly interpretable economic causal effect β .⁶⁰ This parameter is the mean gain of going from $j - 1$ to j for a set of people who pass through q_j even though they may end up far above q_j and start far below q_j . It may differ across j because the distributions of the returns (β) are different across conditioning sets or because of nonlinearity in the returns.

⁵⁷In particular, $(Y_0, \dots, Y_{\bar{s}}) \perp\!\!\!\perp \mathbf{Z}$.

⁵⁸See, e.g., Angrist and Pischke (2009).

⁵⁹We use a scalar Y only to conform to the empirical literature.

⁶⁰Referring to the version of our equation (21) that is in their paper, Angrist and Imbens write “[The Mincer equation] is a structural relationship derived from assumptions about human behavior, but it is not necessarily a causal relationship in the Rubin (1974) sense.” (Angrist and Imbens, 1995, p. 433). Expression (41) is closely related to an expression for the causal effect of schooling reported in Angrist and Imbens (1995), who in our notation, assume that $q_j - q_{j-1} = 1$.

To interpret the Wald estimand using the ordered model of this paper we use the counterfactual notation of Imbens and Angrist (1994), where $D_l(z_k)$ is the random variable D_l when $\mathbf{Z}$ is fixed at z_k . In this expression, assuming that $S(z_2) > S(z_1)$, the numerator of (41) is

$$\begin{aligned} E \sum_{j=1}^{\bar{s}} (Y_j - Y_{j-1}) \sum_{l=j}^{\bar{s}} (D_l(z_2) - D_l(z_1)) &= \sum_{j=1}^{\bar{s}} \sum_{l=j}^{\bar{s}} \{ E(Y_j - Y_{j-1} | S(z_2) = q_l, S(z_1) < q_j) \cdot \Pr(S_l(z_2) = q_l | S(z_1) < q_j) \cdot \Pr(S(z_1) < q_j) \\ &= \sum_{j=1}^{\bar{s}} \sum_{l=j}^{\bar{s}} \sum_{k=0}^{j-1} [E(Y_j - Y_{j-1} | S(z_2) = q_l, S(z_1) = q_k)] \cdot \Pr(S(z_2) = q_l, S(z_1) = q_k) \end{aligned}$$

The term in braces in the immediately preceding expression is the mean change in the outcome of going from $j-1$ to j for persons who at $\mathbf{Z} = z_2$ would stop at q_l years if schooling ($l \geq j$) but who at $\mathbf{Z} = z_1$ would stop at q_k years of schooling ($k \leq j-1$). The final term on the righthand side of this expression is the proportion of this population in the indicated category.

The probability term in the expression can be identified nonparametrically. However, IV does not separately identify the terms in braces. Our structural approach allows us estimate each of the components in this expression and their dependence on the entire $\mathbf{Z}$ process to interpret IV.

5.4 Characterizing What IV Identifies

We now consider what IV estimates in an ordered model when Z^n is a continuous variable. For an analysis of what IV estimates in an unordered model, see Heckman and Urzúa (2010). For simplicity, we continue to consider the case where Y is scalar. The IV estimator is

$$IV(Z^n) = \frac{COV(Y, Z^n)}{COV(S, Z^n)} \cdot \mathbf{1}(COV(S, Z^n) \neq 0). \quad (42)$$

We assume that Z^n is centered around its mean. First consider the numerator of this

expression. From $Y = \sum_{j=0}^{\bar{s}} D_j Y_j$,

$$COV(Y, Z^n) = E \left(\sum_{j=0}^{\bar{s}} D_j Z^n Y_j \right). \quad (43)$$

Consider the j th term and observe that $D_j = \mathbf{I}(\varphi_j(\mathbf{z}) \leq \eta_j, \varphi_{j-1}(\mathbf{z}) > \eta_{j-1}, \dots, \varphi_0(\mathbf{z}) > \eta_0)$, $j \in \{0, \dots, \bar{s}-1\}$, and $D_{\bar{s}} = \mathbf{I}(\varphi_{\bar{s}-1}(\mathbf{z}) > \eta_{\bar{s}-1}, \dots, \varphi_0(\mathbf{z}) > \eta_0)$.

We assume that vector $(Z^n, \varphi_0(\mathbf{Z}), \dots, \varphi_{\bar{s}}(\mathbf{Z}))$ is absolutely continuous with respect to Lebesgue measure with density $f_{Z^n, \varphi_0(\mathbf{Z}), \dots, \varphi_{\bar{s}}(\mathbf{Z})}(z^n, \varphi_0(\mathbf{z}), \dots, \varphi_{\bar{s}}(\mathbf{z}))$ and we further

postulate that Y_j is absolutely continuous with respect to Lebesgue measure conditional on $\mathbf{Z}$, $\varphi_0(\mathbf{Z})$, ..., $\varphi_j(\mathbf{Z})$. Define the support of $\varphi_0(\mathbf{Z}) \times \dots \times \varphi_j(\mathbf{Z})$ as $\Psi_j(\mathbf{Z})$.

Then for $j \in \{1, \dots, s-1\}$

$$E(D_j Z^n Y_j) = \int_{\mathcal{Y}_j} \int_{\mathcal{Z}^n} \int_{\Psi_j(\mathbf{z})} \int_{\underline{\eta}_0}^{\phi_0(\mathbf{z})} \dots \int_{\phi_j(\mathbf{z})}^{\bar{\eta}_j} y_j z^n f_{y_j, \eta_0, \dots, \eta_j, Z^n, \phi_0(\mathbf{z}), \dots, \phi_j(\mathbf{z})}(y_j, z^n, \eta_0, \dots, \eta_j, \phi_0(\mathbf{z}), \dots, \phi_j(\mathbf{z})) dy_j d\eta_0, \dots, d\eta_j dz^n d\phi_0(\mathbf{z}) \dots d\phi_j(\mathbf{z}) \quad (44)$$

where $\mathcal{Z}^n$ is the support of Z^n . For $j=0$, the expression is

$$\int_{\mathcal{Y}_0} y_0 \int_{\mathcal{Z}^n} z^n \int_{\Psi_0(\mathbf{z})} \int_{\phi_0(\mathbf{z})}^{\bar{\eta}_0} f_{y_0, z^n, \eta_0}(y_0, z^n, \phi_0(\mathbf{z}), \eta_0) dy_0 dz^n d\eta_0. \quad (45)$$

Recalling that $Y_l \perp \mathbf{Z}$, $l \in \{0, \dots, s\}$ and $\mathbf{Z} \perp \boldsymbol{\eta}$ and interchanging $\varphi_j(\mathbf{Z})$ and η_j integrating by parts invoking Fubini's theorem, it is as follows that

$$= \int_{\underline{\eta}_j}^{\bar{\eta}_j} \left\{ \int_{\Psi_j(\mathbf{z})} \left[\overbrace{E(Y_j | \eta_j, \phi_{j-1}(\mathbf{z}) \geq \eta_{j-1}, \dots, \phi_0(\mathbf{z}) \geq \eta_0)}^{\text{MTE-like parameter}} \right] \cdot f(\eta_j | \phi_{j-1}(\mathbf{z}) \geq \eta_{j-1}, \dots, \phi_0(\mathbf{z}) \geq \eta_0) \cdot Pr(\phi_{j-1}(\mathbf{z}) \geq \eta_{j-1}, \dots, \phi_0(\mathbf{z}) \geq \eta_0) \right. \\ \left. \cdot \int_{\mathcal{Z}^n} z^n f_{Z^n, \phi_j(\mathbf{z}), \dots, \phi_0(\mathbf{z})}(z^n, \phi_j(\mathbf{z}), \dots, \phi_0(\mathbf{z})) dz^n d\phi_0(\mathbf{z}) \dots d\phi_{j-1}(\mathbf{z}) \right\} d\eta_j \quad (46)$$

where

$$E(Y_j | \eta_j, \phi_{j-1}(\mathbf{z}) \geq \eta_{j-1}, \dots, \phi_0(\mathbf{z}) \geq \eta_0)$$

plays a role analogous to each component of the MTE in the analysis of Heckman and Vytlacil (1999, 2005). For $j=s$, the corresponding expression for the analogous MTE component is

$$E(Y_s | \eta_{s-1}, \phi_{s-2}(\mathbf{z}) \geq \eta_{s-2}, \dots, \phi_0(\mathbf{z}) \geq \eta_0) \quad (47)$$

and $E(D_s Z^n Y_s)$ is defined in an analogous expression.

The denominator of the expression for IV is obtained by a similar argument using

$$E(SZ^n) = \sum_{j=0}^{\bar{s}} q_j E(Z^n D_j). \quad (48)$$

Focusing on the j th component we investigate

$$E(Z^n D_j) = E(Z^n \mathbf{1}(\phi_j(\mathbf{z}) < \eta_j, \phi_{j-1}(\mathbf{z}) \geq \eta_{j-1}, \dots)) \quad (49)$$

$$= \int \int_{\mathcal{Z}^n \Psi_j(\mathbf{z})} \left\{ \int_{\phi_j(\mathbf{z})}^{\bar{\eta}_j} \int_{\eta_{j-1}}^{\phi_{j-1}(\mathbf{z})} \dots \int_{\eta_0}^{\phi_0(\mathbf{z})} z^n \Pr(\phi_j(\mathbf{z}) \leq \eta_j, \phi_{j-1}(\mathbf{z}) \geq \eta_{j-1}, \dots, \phi_0(\mathbf{z}) \geq \eta_0) \right. \\ \left. \cdot f_{Z^n, \phi_j(\mathbf{z}), \dots, \phi_0(\mathbf{z})}(Z^n, \phi_j(\mathbf{z}), \dots, \phi_0(\mathbf{z})) d\eta_j \dots d\eta_0 \right\} dz^n d\phi_j(\mathbf{z}) \dots d\phi_0(\mathbf{z}), \quad (50)$$

$j = \{1, \dots, \bar{s} - 1\}$, with an obvious modification to produce the s^{th} term, and the term for $j = 0$. Because of the dynamics of the model, the simplicity of the MTE framework of Heckman and Vytlacil (1999, 2005) vanishes.

6 Empirical Estimates

This section illustrates the power of our methodology for understanding the causal benefits of education on earnings. As we have just shown, expressions for interpreting what IV estimates can be very complex. But with our model it is possible to generate all of the treatment effects discussed in Section 3 for both ordered and unordered cases.

The standard approach to identifying the causal effect of education on earnings relies on equation (21) that imposes linearity of log earnings in terms of total number of years of schooling (Card, 1999). It ignores the continuation value of schooling (see the discussion in Heckman, Lochner, and Todd, 2006). With our approach, we can investigate returns by educational level and identify the empirical importance of *ex-post* continuation values. We estimate the general ordered model.

We use data for a sample of males from the NLSY79 that is described in detail in Web Appendix B. We do not exploit the full generality of our approach in our parameterization of estimating equations. We use linear-in-parameters specifications of the estimating equations and mixtures of normals representations to approximate the distributions of the unobservables. We use the identification criteria of Williams (2012) to characterize the measurement equations used in this paper. See Web Appendix C.1 for details.

We show that (a) there are substantial but heterogeneous benefits to graduating from high school, though not to GED certification; (b) some people benefit from enrolling in and graduating from college and there is selection on these gains; (c) continuation values—largely neglected in empirical research on the returns to schooling—are empirically substantial;⁶¹ (d) the estimated local average treatment effects are different from the economically interpretable average marginal treatment effects; (e) standard assumptions

⁶¹ Altonji (1993) and Cameron and Heckman (1993) present OLS-based estimates of continuation values of schooling.

imposed in the dynamic discrete choice literature that publicly available future costs affect current decisions are rejected in our data; (f) in the ordered case, the Wald estimator is a weighted combination of gains from heterogeneous subpopulations affected across multiple decisions; and (g) a considerable fraction of the population would continue on to obtain further education if forced to continue past their terminal schooling level.

6.1 Returns To Educational Choices

Table 1 displays the average treatment effect, average marginal treatment effect, average effect of treatment on the treated, and average effect of treatment on the untreated for log wages at age 30 for each educational transition. There is little evidence of selection on gains in the first two educational transitions, but there is a substantial selection on later transitions.^{62,63} We find no statistically significant benefits to GED certification.

The local average treatment effect (LATE) is commonly said to estimate “*the effect*” of schooling for individuals near the margin of indifference for a particular educational choice. The last two rows of Table 1 report the local average treatment effect estimated by two-stage least squares (IV) and the average marginal treatment effect (AMTE). The AMTE and IV estimates are quite different. This result may be due in part to the weakness of our instruments. For some of the transitions the instruments for the nodes associated with the decision to graduate from high school and the decision to graduate from college are not statistically significant in the first stage, resulting in large standard errors and unstable point estimates (See Heckman, Humphries, and Veramendi 2015). In Web Appendix G, we conduct a simulation experiment that documents the poor performance of IV in identifying economically interpretable treatment parameters even when instruments are manufactured to be strong.

The node associated with the choice to enroll in college has two statistically significant instruments – local tuition costs and the presence of a four-year college in the county. For estimates at this node, the IV estimates are statistically significant and substantially larger than the AMTE estimates (as well as the other treatment effects). While it is common in the applied literature to claim that LATE is estimating an economically interpretable parameter like AMTE, this is certainly not the case in this example. This supports our claim that in the context of our dynamic schooling model, IV estimates a difficult-to-interpret treatment effect of little economic or policy relevance.

6.2 Continuation Values

One benefit of schooling is access to further schooling.⁶⁴ Specifically, the choice to graduate from high school gives the option to go to college and the choice to enroll in college gives the option of attaining a four-year college degree. The direct effect of an educational choice is the direct gain from choosing to attain the next level of schooling while the continuation value of an educational choice is the probability of continuing education times the benefits

⁶²We test this by comparing the differences between treatment on the treated and treatment on the untreated. Under the null of no selection bias, these parameters should be the same.

⁶³For details on the construction of treatment effects see Section 3. We report dynamic treatment effects as defined by Equation (11).

⁶⁴See Weisbrod (1962) and Comay, Melnik, and Pollatschek (1973).

of that additional education. For example, the continuation value for enrolling in college is the probability of attaining a four-year degree times the benefit of the degree. For high-ability individuals, the benefits of college are large, and the probability of attending is close to 1. For such individuals, the continuation value constitutes the bulk of the return to graduating from college. For others, the continuation value is much lower. Table 1 shows the total effect and the direct effect, and the percentage of the total effect arising from continuation values.

The continuation value is small for those who are indifferent between graduating from high school. The direct effect and the total effect are nearly the same. On the other hand, the continuation value is an important component of the average effects (ATE) and is especially important for those choosing to graduate from high school (TT). A similar pattern holds for the college enrollment decision, where the continuation value is larger for those who choose to enroll than those who are indifferent or decide not to enroll.

6.3 Are Agents Forward-Looking?

When publicly available variables that are estimated to be empirically important in determining later educational choices are used as possible determinants of earlier educational choices, they are all found to be statistically insignificant. For example, two instruments that are found to be statistically significant in determining the college enrollment decision – local college tuition and the presence of a four-year college in the county – are found to be statistically insignificant in determining the decision to graduate from high school.⁶⁵ This evidence is consistent either with the claim that agents are not forward-looking or that agent abilities to process publicly available information are weak.⁶⁶ The implicit assumption in many Bellman-equation-based models of educational decisions and their consequences is thus called into question.

6.4 Introducing Colleges into Localities: Interpreting What IV Estimates

This section uses simulations from our model to decompose LATE into its underlying components, as presented in Section 5.3 for the ordered model. The instrument that we use (availability of college) only affects college enrollment decisions and not high school enrollment decisions. So it is meaningful to apply the ordered model for the analysis we report here because it characterizes transitions beyond high school graduation. Let $S(z)$ denote the final level of schooling selected when $Z = z$. As before, q_j is the years of schooling associated with attainment at node j . We consider a binary valued instrument ($Z \in \{z_1, z_2\}$) and assume that $S(z_2) > S(z_1)$.

As shown in Section 5.3, for the ordered case, $LATE(z_1, z_2)$ includes the mean gain of going from $j-1$ to j for people who pass through j even though they may end up far above j and start far below it. It may differ across j because the distributions of the returns are different across conditioning sets and because of the intrinsic nonlinearity in the returns.

⁶⁵See Heckman, Humphries, and Veramendi (2015) for details.

⁶⁶Taber (2000) reports that for his estimated model, the presence of college in the vicinity has no effect on *college enrollment* and has a similar discussion why the estimated effect is weak. Unlike our analysis, he *constrains* the effect of having a college in the vicinity on high school graduation to be zero.

Consider the college-in-county indicator used in the IV analysis of Kling (2001). Let $Z = z_1$ when no college is present in the county and $Z = z_2$ when a college is present.⁶⁷ We assume this instrument affects college enrollment and college graduation decisions.⁶⁸ For a binary instrument that affects only the final two educational choices, three different sub-populations are affected: (1) those who are induced to enroll in college but do not graduate, (2) those induced to enroll in college and who go on to graduate, and (3) those who previously enrolled in college and who would not have graduated who are induced to graduate by the policy.

Simulating our model, we decompose the Wald estimator for the effect of schooling on log wages into the expected gains at each transition for each sub-population as well as the weights $W_j, j = 2, \dots, s$. LATE is not the return to any particular population at any particular margin, but rather a weighted average of the returns to a year of schooling for each of the affected sub-populations across different transitions. It weights the expected gains differentially and depends on which sub-populations are induced to change schooling levels. Even if an instrument affects only one margin, individuals induced to change at that margin may still go on to complete additional education, again making LATE a weighted average of returns. Table 2 shows how the returns differ by margin and sub-population. The subpopulations induced to switch by the change in the instrument differ substantially by observable characteristics and in their average levels of cognitive and non-cognitive endowments.

Table 3 shows how the average levels of cognitive and socio-emotional endowments and other background characteristics differ across the sub-populations induced to change final schooling levels by the college introduction instrument. The columns show the three different populations induced to switch by changing Z .⁶⁹

6.5 Regret

Some individuals continue their education even though the expected *ex-post* benefit is negative. This behavior is not ruled out in our model as the agent decision rules are not required to be based on realized gains. From our model, an estimated 18% of high school dropouts would have enrolled in college if they had graduated from high school (and gotten the cost shock associated with completing high school) and 4% would have gone on to complete four-year degrees. Similarly, if terminal high school graduates were forced to enroll in college, an estimated 28% would then choose to complete college.⁷⁰ This analysis suggests that *ex-post* many agents regret their early educational decisions.⁷¹

⁶⁷Oreopoulos and Salvanes (2011) present a summary of papers using presence of a local college in a county as an instrument.

⁶⁸See Section 6.3 for discussion on why we do not allow the instrument to affect the high school graduation decision.

⁶⁹The presence of a college in the vicinity does not affect the high school graduation decision (see Table F.1 of the Web Appendix for details). In Section G of the Web Appendix, we create counterfactual simulations where the presence of the local college variable enters into the high school graduation decision with different counterfactual coefficients. Even for a large instrument, IV estimates on the simulated data are far from the true AMTE or ATE.

⁷⁰See Table D.1 in the Web Appendix.

⁷¹Our model does not let us differentiate between behaviors that result in regret and behavior driven by costs not observed by the economist.

7 Summary and Conclusions

This paper develops and estimates a model of treatment effects that arise in multistage decision models. It gives economic content to the treatment effect literature without imposing the sometimes controversial rationality and functional form assumptions standard in the literature on dynamic discrete choice.

Identification of our model for both ordered and unordered versions of the model can be obtained from multiple sources: (a) instrumental variables, as is standard in the treatment effect literature, (b) through conditional independence assumptions, as is conventional in the dynamic discrete choice literature and also in the literature on matching, or (c) through combinations of these assumptions.

Our analysis expands the tool kit available to the applied economist. We define new dynamic treatment effects and decompose them into direct effects of forcing agents to move or not to move to the immediate next stage of a dynamic decision process and the effects of continuing beyond the next stage. Our *ex-post* continuation values approximate the *ex-ante* continuation values of a structural dynamic discrete choice model. Using our framework we interpret what IV estimates in a dynamic multistage model. Due to the sequential nature of agent decision making, instruments that affect only one argument of a multistage decision model affect treatment effects and estimated transition probabilities of all subsequent stages. We clarify the crucial role of the implicit assumption of an ordered choice model that underlies the LATE-inspired IV literature.

We estimate a model of educational choices on earnings for a sample of American males, and use the estimated model to generate a variety of economically interpretable treatment effects and to compare them with the treatment effects reported in the statistics literature. We find sharp discrepancies between the treatment effects generated from the two approaches. In our samples, LATE is a poor approximation of the economically interpretable average marginal treatment effect. We use our model to decompose the Wald estimator for a dynamic discrete choice model into economically interpretable components and weights on those components.

Contrary to a behavioral assumption maintained in the dynamic discrete literature, we test for and do not find evidence of forward-looking behavior. We find that *ex-post* a large fraction of agents would continue on to further education if forced to make transitions at earlier stages.

Web Appendix

Refer to Web version on PubMed Central for supplementary material.

References

- Abbring, JH.; Heckman, JJ. Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation. In: Heckman, J.; Leamer, E., editors. Handbook of Econometrics. Vol. 6B. Elsevier; Amsterdam: 2007. p. 5145-5303.

- Adda, J.; Cooper, RW. *Dynamic Economics: Quantitative Methods and Applications*. The MIT Press; Cambridge, Mass: 2003.
- Aguirregabiria V. Another look at the identification of dynamic discrete decision processes: An application to retirement behavior. *Journal of Business & Economic Statistics*. 2010; 28(2):201–218.
- Almlund, M.; Duckworth, A.; Heckman, JJ.; Kautz, T. *Personality Psychology and Economics*. In: Hanushek, EA.; Machin, S.; Wößmann, L., editors. *Handbook of the Economics of Education*. Vol. 4. Elsevier; Amsterdam: 2011. p. 1-181.
- Altonji JG. The Demand for and Return to Education When Education Outcomes are Uncertain. *Journal of Labor Economics*. 1993; 11(1):48–83.
- Anderson, TW.; Rubin, H. *Statistical Inference in Factor Analysis*. In: Neyman, J., editor. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 5. University of California Press; Berkeley: 1956. p. 111-150.
- Angrist JD, Imbens GW. Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association*. 1995; 90(430): 431–442.
- Angrist, JD.; Pischke, J-S. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press; Princeton: 2009.
- Angrist JD, Pischke J-S. The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*. 2010; 24(2):3–30.
- Arcidiacono P, Ellickson PB. Practical Methods for Estimation of Dynamic Discrete Choice Models. *Annual Review of Economics*. 2011; 3:363–394.
- Bartlett MS. The Statistical Conception of Mental Factors. *British Journal of Psychology*. 1937; 28(1): 97–104.
- Blevins JR. Nonparametric Identification of Dynamic Decision Processes with Discrete and Continuous Choices. *Quantitative Economics*. 2014; 5:531–554.
- Bonhomme S, Robin J-M. Consistent Noisy Independent Component Analysis. *Journal of Econometrics*. 2009; 149:12–25.
- Borghans L, Duckworth AL, Heckman JJ, ter Weel B. The Economics and Psychology of Personality Traits. *Journal of Human Resources*. 2008; 43(4):972–1059.
- Cameron, SV.; Heckman, JJ. Unpublished working paper. University of Chicago, Department of Economics; 1987. Son of CTM: The DCPA Approach Based on Discrete and Factor Structure Models.
- Cameron SV, Heckman JJ. The Nonequivalence of High School Equivalents. *Journal of Labor Economics*. 1993; 1(1 Part 1):1–47.
- Cameron SV, Heckman JJ. Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males. *Journal of Political Economy*. 1998; 106(2):262–333.
- Cameron SV, Heckman JJ. The Dynamics of Educational Attainment for Black, Hispanic, and White Males. *Journal of Political Economy*. 2001; 109(3):455–499.
- Card, D. The Causal Effect of Education on Earnings. In: Ashenfelter, O.; Card, D., editors. *Handbook of Labor Economics*. Vol. 5. North-Holland; New York: 1999. p. 1801-1863.
- Carneiro P, Hansen K, Heckman JJ. Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice. *International Economic Review*. 2003; 44(2):361–422.
- Carneiro P, Heckman JJ, Vytlacil EJ. Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin. *Econometrica*. 2010; 78(1):377–394. [PubMed: 20209119]
- Carneiro P, Heckman JJ, Vytlacil EJ. Estimating Marginal Returns to Education. *American Economic Review*. 2011; 101(6):2754–2781. [PubMed: 25110355]
- Chamberlain, G. An Instrumental Variable Interpretation of Identification in Variance Components and MMIC Models. In: Taubman, P., editor. *Kinometrics : Determinants of Socioeconomic Success Within and Between Families*. North-Holland; New York: 1977.

- Comay Y, Melnik A, Pollatschek MA. The Option Value of Education and the Optimal Path for Investment in Human Capital. *International Economic Review*. 1973; 14(2):421–435.
- Cunha F, Heckman JJ, Navarro S. The Identification and Economic Content of Ordered Choice Models with Stochastic Cutoffs. *International Economic Review*. 2007; 48(4):1273–1309.
- Cunha F, Heckman JJ, Schennach SM. Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*. 2010; 78(3):883–931. [PubMed: 20563300]
- Eckstein Z, Wolpin KI. The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey. *Journal of Human Resources*. 1989; 24(4):562–598.
- Eisenhauer P, Heckman J, Mosso S. Estimation of Dynamic Discrete Choice Models by Maximum Likelihood and the Simulated Method of Moments. Forthcoming, *International Economic Review*. 2015
- Eisenhauer P, Heckman JJ, Vytlačil EJ. Generalized Roy Model and Cost-Benefit Analysis of Social Programs. Forthcoming, *Journal of Political Economy*. 2015
- Freyberger, J. Nonparametric panel data models with interactive fixed effects. Department of Economics, University of Wisconsin; Madison: 2014.
- Gensowski, M. Discussion Paper 8235. IZA; 2014. Personality, IQ, and Lifetime Earnings.
- Geweke, J.; Keane, M. Computationally Intensive Methods for Integration in Econometrics. In: Heckman, JJ.; Leamer, E., editors. *Handbook of Econometrics*. Vol. 5. North-Holland; New York: 2001. p. 3463-3568.
- Haavelmo T. The Statistical Implications of a System of Simultaneous Equations. *Econometrica*. 1943; 11(1):1–12.
- Hansen LP, Sargent TJ. Formulating and Estimating Dynamic Linear Rational Expectations Models. *Journal of Economic Dynamics and Control*. 1980; 2(1):7–46.
- Heckman, J.; Humphries, JE.; Veramendi, G. The Causal Effects of Education on Earnings and Health. Department of Economics, University of Chicago; 2015. Unpublished manuscript
- Heckman, JJ. Statistical Models for Discrete Panel Data. In: Manski, C.; McFadden, D., editors. *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press; Cambridge, MA: 1981. p. 114-178.
- Heckman JJ. The Principles Underlying Evaluation Estimators with an Application to Matching. *Annales d'Economie et de Statistiques*. 2008; 91–92:9–73.
- Heckman JJ. Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature*. 2010; 48(2):356–398. [PubMed: 21743749]
- Heckman, JJ.; Humphries, JE.; Kautz, T., editors. *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. University of Chicago Press; Chicago: 2014a.
- Heckman, JJ.; Humphries, JE.; Kautz, T. Who Are the GEDs?. In: Heckman, JJ.; Humphries, JE.; Kautz, T., editors. *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. University of Chicago Press; Chicago: 2014b.
- Heckman, JJ.; Kautz, T. Fostering and Measuring Skills: Interventions That Improve Character and Cognition. In: Heckman, JJ.; Humphries, JE.; Kautz, T., editors. *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. University of Chicago Press; Chicago, IL: 2014. p. 341-430.
- Heckman, JJ.; Lochner, LJ.; Todd, PE. Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In: Hanushek, EA.; Welch, F., editors. *Handbook of the Economics of Education*. Vol. 1. Elsevier; Amsterdam: 2006. p. 307-458.chap 7
- Heckman JJ, Navarro S. Dynamic Discrete Choice and Dynamic Treatment Effects. *Journal of Econometrics*. 2007; 136(2):341–396.
- Heckman, JJ.; Pinto, R. Beyond Monotonicity. University of Chicago; 2015a. Unpublished manuscript
- Heckman JJ, Pinto R. Causal Analysis after Haavelmo. *Econometric Theory*. 2015b; 31(1):115–151. [PubMed: 25729123]
- Heckman JJ, Pinto R, Savelyev PA. Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*. 2013; 103(6):1–35. [PubMed: 25506083]

- Heckman, JJ.; Schennach, SM.; Williams, B. Matching on Proxy Variables. University of Chicago, Department of Economics; 2011. Revised 2014
- Heckman JJ, Smith JA, Clements N. Making the Most Out Of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts. *Review of Economic Studies*. 1997; 64(4):487–535.
- Heckman JJ, Urzúa S. Comparing IV With Structural Models: What Simple IV Can and Cannot Identify. *Journal of Econometrics*. 2010; 156(1):27–37. [PubMed: 20440375]
- Heckman JJ, Urzúa S, Vytlačil EJ. Understanding Instrumental Variables in Models with Essential Heterogeneity. *Review of Economics and Statistics*. 2006; 88(3):389–432.
- Heckman JJ, Urzúa S, Vytlačil EJ. Instrumental Variables in Models with Multiple Outcomes: The General Unordered Case. *Les Annales d'Economie et de Statistique*. 2008; 91–92:151–174.
- Heckman JJ, Vytlačil EJ. Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Sciences*. 1999; 96(8): 4730–4734.
- Heckman JJ, Vytlačil EJ. Structural Equations, Treatment Effects and Econometric Policy Evaluation. *Econometrica*. 2005; 73(3):669–738.
- Heckman, JJ.; Vytlačil, EJ. Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation. In: Heckman, J.; Leamer, E., editors. *Handbook of Econometrics*. Vol. 6B. Elsevier; Amsterdam: 2007a. p. 4779-4874.
- Heckman, JJ.; Vytlačil, EJ. Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments. In: Heckman, J.; Leamer, E., editors. *Handbook of Econometrics*. Vol. 6B. Elsevier; Amsterdam: 2007b. p. 4875-5143.chap 71
- Hu Y, Schennach SM. Instrumental Variable Treatment of Nonclassical Measurement Error Models. *Econometrica*. 2008; 76(1):195–216.
- Imbens GW, Angrist JD. Identification and Estimation of Local Average Treatment Effects. *Econometrica*. 1994; 62(2):467–475.
- Kautz, T.; Zannoni, W. Measuring and Fostering Non-Cognitive Skills in Adolescents: Evidence from Chicago Public Schools and the OneGoal Program. University of Chicago, Department of Economics; 2015. Unpublished manuscript
- Keane, MP.; Todd, PE.; Wolpin, KI. The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications, vol. 4 of *Handbook of Labor Economics*. Elsevier; 2011.
- Keane MP, Wolpin KI. The Career Decisions of Young Men. *Journal of Political Economy*. 1997; 105(3):473–522.
- Kling JR. Interpreting Instrumental Variables Estimates of the Returns to Schooling. *Journal of Business and Economic Statistics*. 2001; 19(3):358–364.
- Laibson, D. Intertemporal Decision Making. In: Nadel, L., editor. *Encyclopedia of Cognitive Science*. Nature Publishing Group; London: 2003.
- Lewbel A. Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables. *Journal of Econometrics*. 2000; 97(1):145–177.
- Lochner, L. Nonproduction Benefits of Education: Crime, Health, and Good Citizenship. In: Hanushek, EA.; Machin, S.; Woessmann, L., editors. *Handbook of the Economics of Education*. Vol. 4. Amsterdam: North-Holland; 2011. p. 183-282.
- Madansky A. Instrumental Variables in Factor Analysis. *Psychometrika*. 1964; 29(2):105–113.
- Magnac T, Thesmar D. Identifying Dynamic Discrete Decision Processes. *Econometrica*. 2002; 70(2): 801–816.
- Matzkin, RL. Restrictions of Economic Theory in Nonparametric Methods. In: Engle, R.; McFadden, D., editors. *Handbook of Econometrics*. Vol. 4. North-Holland; New York: 1994. p. 2523-2558.
- McMahon, WW. The Total Return to Higher Education: Is There Underinvestment for Development?. University of Illinois; Urbana-Champaign: 2014.
- Mincer, J. Schooling, Experience and Earnings. Columbia University Press for National Bureau of Economic Research; New York: 1974.

- Murphy SA. Optimal Dynamic Treatment Regimes. *Journal of the Royal Statistical Association, Series B*. 2003; 65(2):331–355.
- Oreopoulos P, Salvanes KG. Priceless: The Nonpecuniary Benefits of Schooling. *Journal of Economic Perspectives*. 2011; 25(1):159–184.
- Pagan, A.; Ullah, A. *Nonparametric Econometrics*, no. 9780521355643 in Cambridge Books. Cambridge University Press; 1999.
- Pudney SE. Estimating Latent Variable Systems When Specification is Uncertain: Generalized Component Analysis and the Eliminant Method. *Journal of the American Statistical Association*. 1982; 77(380):883–889.
- Quandt RE. The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association*. 1958; 53(284):873–880.
- Quandt RE. A New Approach to Estimating Switching Regressions. *Journal of the American Statistical Association*. 1972; 67(338):306–310.
- Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. 1983; 70(1):41–55.
- Rust, J. Structural Estimation of Markov Decision Processes. In: Engle, R.; McFadden, D., editors. *Handbook of Econometrics, Volume*. North-Holland; New York: 1994. p. 3081–3143.
- Schennach, S. Measurement Error in Nonlinear Models A Review. In: Acemoglu, D.; Arellano, M.; Dekel, E., editors. *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*. Vol. 3. Cambridge University Press; New York: 2013.
- Schennach S, White H, Chalak K. Local indirect least squares and average marginal effects in nonseparable structural systems. *Journal of Econometrics*. 2012; 166(2):282–302.
- Skrondal A, Laake P. Regression among Factor Scores. *Psychometrika*. 2001; 66(4):563–576.
- Steinberg, L. *Age of Opportunity: Lessons from the New Science of Adolescence*. Eamon Dolan/Houghton Mifflin Harcourt; New York: 2014.
- Taber CR. Semiparametric Identification and Heterogeneity in Discrete Choice Dynamic Programming Models. *Journal of Econometrics*. 2000; 96(2):201–229.
- Vytlacil EJ. Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica*. 2002; 70(1):331–341.
- Vytlacil EJ. A Note on Additive Separability and Latent Index Models of Binary Choice: Representation Results. *Oxford Bulletin of Economics and Statistics*. 2006a; 68(4):515–518.
- Vytlacil EJ. Ordered Discrete-Choice Selection Models and Local Average Treatment Effect Assumptions: Equivalence, Nonequivalence, and Representation Results. *Review of Economics and Statistics*. 2006b; 88(3):578–581.
- Weisbrod BA. Education and Investment in Human Capital. *Journal of Political Economy*. 1962; 70:106–123. 5, Part 2: Investment in Human Beings.
- White H, Chalak K. Settable Systems: An Extension of Pearl’s Causal Model with Optimization, Equilibrium, and Learning. *Journal of Machine Learning Research*. 2009; 10:1759–1799.
- Williams, B. Identification of Factor Models. Department of Economics, University of Chicago; 2012.

Figure 1.
An Ordered Multistage Dynamic Decision Model

Figure 2.
Unordered Dynamic Model

Table 1

The Effects of Education on Log Wages (age 30)

Estimate	Graduate High School		Enroll in College		Graduate College		Earn GED	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
ATE	0.094 *	(0.056)	0.134 **	(0.025)	0.114 **	(0.037)	0.059	(0.052)
(Dir)	0.036	(0.056)	0.085 **	(0.029)				
(% Cont Val.)	73%		25%					
TT	0.093	(0.072)	0.140 **	(0.031)	0.141 **	(0.042)	0.063	(0.053)
(Dir)	0.021	(0.068)	0.062	(0.040)				
(% Cont Val.)	77%		56%					
TUT	0.100 **	(0.029)	0.128 **	(0.026)	0.079 **	(0.036)	0.055	(0.055)
(Dir)	0.089 **	(0.031)	0.109 **	(0.030)				
(% Cont Val.)	11%		15%					
AMTE	0.093 **	(0.028)	0.101 **	(0.023)	0.110 **	(0.034)	0.057	(0.051)
(Dir)	0.087 **	(0.032)	0.077 **	(0.028)				
(% Cont Val.)	7%		24%					
IV	-0.763	(0.807)	0.511 *	(0.207)	0.025	(0.852)	0.130	(0.551)

Notes: Standard errors (in parenthesis) and significance levels

* p<0.05,

** p<0.01.

All standard errors calculated with 200 bootstrap samples. Each column presents the average effect of an educational decision. Total effects are reported on top of the “direct” component. Each schooling level provides the option to pursue higher schooling levels, while final schooling levels do not provide such an option. “DIR” represents the direct effect of attaining the schooling level and stopping there. The direct effect is the same as the total effect for terminal nodes. “% Cont Val.” reports the percentage of the total effect that is generated by access to future educational decisions and resulting continuation values. The TT row presents the average effect for those who chose a higher level of schooling ($D_j = 0$), and TUT presents the average effect for those who do not choose the next stage of schooling ($D_j = 1$). AMTE presents the average effect for those who are indifferent between choosing a higher level of schooling or not. The table also presents the estimated treatment effect from two stage least squares controlling for family background variables, factors, and decision-specific instruments by educational choice (“IV”). For graduating from high school the instruments are long run local unemployment rate and current local unemployment rate at 17. The choice to earn a GED additionally includes, as an instrument, the state-wide passing difficulty of the GED. The choice to enroll in college includes long run local unemployment rate, current local unemployment rate at 17, presence of a 4-year college in the county, and local college tuition. The choice to graduate from college includes local long-run unemployment at age 22, current local unemployment at age 22, and local college tuition at age 22. The IV models include background controls and factor scores. For exact formulations of the treatment effects, see Section 3. The difference between TT and TUT is statistically significantly different from 0 at the 5% level for the decision to graduate from college.

Table 2

Decomposing the LATE Estimator of the Effect of Education on Log Wages

Group	Return	Weight	q_j	Avg Yearly Return
$S(z_1) = 1, S(z_2) = 2$	$E[Y_2 - Y_1] = 0.087$	$W_{1,2} = 0.289$	1.790	0.049
$S(z_1) = 1, S(z_2) = 3$	$E[Y_2 - Y_1] = 0.084$	$W_{1,3} = 0.204$	1.790	0.047
$S(z_1) = 1, S(z_2) = 3$	$E[Y_3 - Y_2] = 0.094$	$W_{1,3} = 0.341$	2.988	0.031
$S(z_1) = 2, S(z_2) = 3$	$E[Y_3 - Y_2] = 0.119$	$W_{2,3} = 0.166$	2.988	0.040
Total=1			$\rho_{LATE} = 0.041$	

Notes: The LATE estimator is the weighted sum of returns. The “Normalized Return”, is the average return per year of schooling for that specific sub population $\rho_j = \frac{E[Y_j - Y_{j-1}]}{(q_j - q_{j-1})}$. The Return column shows the expected return for the specific group defined in the Group column. q_j is the difference in the average number of years of schooling for individuals transiting from $j-1$ to j . Of those induced to change schooling levels by the instrument, 48.7% moved from high school graduate to some college, 34.5% moved from high school graduate to college graduate, and 16.8% moved from some college to college graduate.

Table 3

Means of Observables and Endowments for Groups that Constitute the LATE Decomposition

	Full Sample	HS to Some Coll (1 → 2)	HS to Coll Grad (1 → 3)	Some Coll to Coll Grad (2 → 3)
Cognitive	0.001 (0.679)	−0.006 (0.605)	0.286 (0.579)	0.321 (0.570)
Non-Cognitive	0.001 (0.652)	0.032 (0.596)	0.249 (0.566)	0.229 (0.567)
Black	0.117 (0.322)	0.116 (0.320)	0.083 (0.275)	0.096 (0.295)
Hispanic	0.067 (0.250)	0.067 (0.250)	0.044 (0.204)	0.057 (0.231)
Broken Home	0.241 (0.428)	0.235 (0.424)	0.157 (0.364)	0.197 (0.397)

Notes: Mean coefficients; sd in parentheses*
 $p < 0.10$,**
 $p < 0.05$,***
 $p < 0.01$